Understanding Gross Worker Flows Across U.S. States

Daniele Coen-Pirani*
Tepper School of Business
Carnegie Mellon University

This Version: January 2008. First Version: February 2006.

Abstract

A surprising but robust characteristic of workers' migration patterns across locations (states and metropolitan areas) within the U.S. is the positive correlation between inflow and outflow rates. This pattern cannot be accounted for by standard equilibrium models of employment reallocation across geographic areas in which net and gross flows of workers coincide. Further, micro-level evidence shows that inflows and outflows of workers tend to simultaneously occur within narrowly defined demographic groups, suggesting that the positive inflow-outflow correlation is not the symptom of a changing demographic composition of employment across locations. This paper develops and estimates a dynamic general equilibrium model of gross and net migration flows to explain this pattern. Due to a selection effect, workers migrating into a location have a higher propensity to migrate again than workers already living there. Thus, U.S. states that absorb large numbers of internal migrants also tend to display relatively large outflow rates. The time-series pattern of inflow and outflow rates across states is consistent with this interpretation.

Keywords: Migration, States, Employment, Gross Flows, Net Flows, Wages, Rents, Productivity, Land.

JEL Classification: J0, J6, E1, R0.

*E-mail: coenp@andrew.cmu.edu. Thanks, without implication, to seminar participants at Carnegie Mellon, U. of Texas-Austin, SUNY Albany, Cleveland Fed, UQAM, LABORatorio Revelli, the 2006 Midwest Economics Association Meetings, the 2006 Society for Economic Dynamics Meetings, the 2007 Texas Monetary Conference, the 2007 Midwest Macro Meetings, the 2007 Econometrics Society North American Summer Meetings, and to Nathan Balke, Yoonsoo Lee, Chris Flinn, Holger Sieg, Bruno Contini, Fallaw Sowell, Marla Ripoll, Maria Ferreyra, Dennis Epple, Alexis Leon, Iourii Manovski, Rui Castro, B. Ravikumar, Rodi Manuelli, Ted Temzelides for useful discussions and suggestions. Richard Lowery provided able research assistance.

1 Introduction

A surprising but robust characteristic of workers' migration patterns across locations within the U.S. is the positive correlation between inflow and outflow rates. In other words, locations that attract large numbers of internal migrants relative to their population also tend to experience relative large out-migration in relative terms. Figure 1 plots gross outflow rates of workers from U.S. states against gross inflow rates using pooled data from the 1970-2000 U.S. Censuses. Controlling for year effects, the correlation coefficient between these two variables is 0.56 and statistically significant. A similar relationship holds also across metropolitan areas. For example, the cross-sectional correlation between inflow and outflow rates of workers across the 100 largest U.S. metropolitan areas in 2000 is 0.41 and also statistically significant.[1]

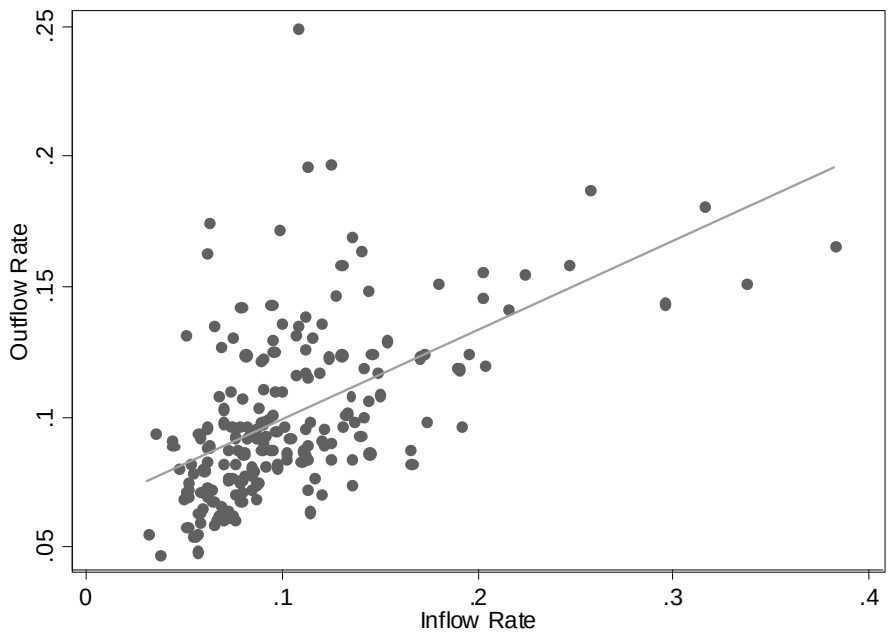

Figure 1: Scatter plot of inflow and outflow rates of workers across U.S. states, 1970-2000. The correlation coefficient between inflow and outflow rates is 0.56 and has been computed using year fixed-effects and weighting each state by its relative population in the relevant Census year. The figure also reports a regression line. Data source: my computations based on data from the U.S. Census of Population and Housing.

The positive correlation between inflow and outflow rates cannot be accounted for by standard models of employment reallocation across geographic areas in which net and gross

[1] Both correlation coefficients were computed by weighting each location by its relative workforce. The p-values associated with the correlation coefficients are both smaller than 0.0001. Worker flows are constructed using information on the worker's state (or metropolitan area) of residence at the time of the Census and her state (or metropolitan area) of residence 5 years before the Census year. Section 2 and Appendix A describe the data and provide details on how inflow and outflow rates are computed.

flows of workers coincide (e.g. Lucas and Prescott, 1974). Less obviously, micro data from the Census reveal that workers moving into a U.S. state tend to be observationally similar to workers migrating out of it. In particular, these workers tend to have the same age and education and to work in the same industries. This evidence runs counter to the argument advanced by Larry Sjaastad (1962, 1961), who in his classic work on migration stated that the "somewhat paradoxical relation between gross in- and out-migration may be substantially an aggregation problem," meaning that "gross migration reflects the degree to which the labor force is being reshaped by changing demand and supply conditions among industries."

This paper explains the positive cross-sectional relationship between inflow and outflow rates using a dynamic general equilibrium model of net and gross migration across locations. The model economy is composed by a set of local labor and land markets. Labor demand shocks drive the net reallocation of workers across markets while both ex-ante and ex-post heterogeneity among workers gives rise to "excess" geographic reallocation of workers. The estimated version of the model accounts well for the cross-sectional and time-series properties of net and gross worker flows and the relationship between worker flows and the local prices of labor and land.

The key features of the model that explain the positive cross-sectional correlation between inflow and outflow rates are a selection effect and the presence of location-specific shocks. The selection effect implies that workers who recently moved to a new location are characterized by a higher propensity to migrate again than workers who had lived there for a longer period of time. The presence of location-specific shocks gives rise to dispersion in net and gross inflows of workers across locations. Putting these two elements together, in the model locations hit by positive shocks attract a relatively large number of internal migrants as a fraction of their population and experience in subsequent periods a relatively high gross outflow rate. Since gross inflow rates are persistent in the model as well as in the data, gross inflow and outflow rates are also contemporaneously positively correlated.

In the model, recent migrants display a higher propensity to migrate again than workers who remained in the same location. This result is due to two sources of heterogeneity among workers. First, following Jovanovic (1979), Flinn (1986) and, more recently, Kennan and Walker (2006), I assume that workers can observe aspects of their match with a location only after having migrated there. Thus, recent migrants will be characterized on average by a relatively lower match with the location than workers already living there. Second, following Farber (1994), I also allow for ex-ante heterogeneity. Some workers are assumed to be characterized by higher mobility rates than other workers. Thus, high-mobility workers will tend to be disproportionately represented in the pool of migrants, also leading to a high rate of out-migration from fast-growing locations.

In addition to being consistent with the key facts about worker flows, labor and land prices, the model generates two additional predictions that are consistent with the selection effect described above. First, the latter implies that states with relatively large gross inflow rates in the current period should experience relatively large gross outflow rates in *future* periods. The *contemporaneous* correlation between inflow and outflow rates obtains because gross inflow rates are persistent over time. This implication of the model can be tested using data on gross worker flows across several Census decades. The results are consistent with the

model's mechanism: in a regression of current outflow rates on both contemporaneous and lagged inflow rates, the estimated coefficient on current inflows is not statistically different from zero, while the coefficient on lagged inflows is positive and significant. Second, the feedback from lagged inflows to current outflows implies that gross inflow rates should be more persistent than net flow rates. The additional persistence of gross inflows is due to the need of a location to replace workers who migrate for idiosyncratic reasons: relative high gross inflow rates lead to relative high gross outflow rates, which, for given process followed by net flows, induce more gross inflows, etc. The higher persistence of gross inflow rates is clearly evident in the Census data.

A novel contribution of the paper is to use micro data from the U.S. Census of Population and Housing to construct measures of net and gross flows of workers across U.S. states and to characterize the main cross-sectional and time-series properties of these flows. Demographic information about workers' education, age, and industry of employment allows me to control for possible composition effects. Similarly, individual-level information on earnings and housing unit rents can be used to construct measures of state and time-specific labor earnings and land rents. The Census data reveals interesting information about the features of migration flows and the nature of the shocks responsible for net employment reallocation at the state-level. First, differences in net flows rates across U.S. states are mainly due to differences in gross inflow rates. Outflow rates tend to be relatively high in locations experiencing large positive as well as large negative net flows. Second, net and gross flows of workers are very persistent across different Census years. Thus, while state-level business cycles might give rise to some of the observed migration flows, much of the observed reallocation of employment across states is due to shocks characterized by much lower frequency than cyclical shocks. Third, in the cross-section, states with relatively high earnings are also characterized by relatively high land prices. This pattern points to the importance of labor demand rather than labor supply shocks in driving geographical reallocation of workers (Roback, 1982). The estimated version of the theoretical model developed in this paper is consistent with these stylized facts.

The rest of the paper is organized as follows. Section 1.1 reviews the related literature. Section 2 describes the data and the stylized facts. Section 3 presents the model. Section 4 describes the estimation of the model. Section 5 discusses the model's fit. Section 6 provides some empirical evidence on the mechanism of the model. Section 7 evaluates the relative importance of ex-ante and ex-post heterogeneity. Section 8 concludes. The appendices offer a more detailed description of the data, of the construction of the flow variables, and of the algorithm used to solve and estimate the model.

1.1 Related Literature

This paper is related to several literatures. The general equilibrium approach to "migration", broadly meant to include moves from one employment status or sector to another, was pioneered by Lucas and Prescott (1974)'s island model of the labor market.[2] In the latter,

[2] Kambourov and Manovskii (2004) use the Lucas-Prescott model to study trends in occupational mobility and wage inequality. Alvarez and Veracierto (2000 and 2006) analyze labor market policies. Alvarez and Shimer (2007) study a continuous-time version of the Lucas-Prescott model to study different types of

net and gross flows of workers coincide and net flows across locations are driven by shocks to local labor demand. The present paper is a version of Lucas and Prescott's model in which also workers are hit by idiosyncratic location-specific shocks, giving rise to gross flows of workers in excess to net flows.

The focus on gross flows links this paper to Jovanovic and Moffitt (1990) who consider an overlapping-generations version of the Lucas-Prescott model to study gross flows of workers across industries. Similarly to this paper, Jovanovic and Moffitt focus on learning about match quality as a mechanism to generate excess reallocation of workers across sectors. In addition to focusing on geographic, as opposed to sectoral, mobility, my paper differs from Jovanovic and Moffitt (1990) in two other important dimensions. First, in their model workers live for only two periods and can therefore move only once in their lifetime. This assumption simplifies the analysis considerably but is ill-suited for empirical analysis. Second, Jovanovic and Moffitt focus on an equilibrium in which sectoral productivity shocks are small enough that gross inflows into each sector are always strictly positive. In this case unit wages are equalized across sectors and gross outflows from sectors occur only because workers realize that the match with their employer is of a relatively low quality. This implies that the Jovanovic-Moffitt model cannot account for the U-shaped relationship between net flow rates and gross outflow rates of workers across U.S. states. In order to account for the latter it is necessary to take explicitly into account the possibility of corner solutions in which voluntary gross inflows into a location are occasionally zero. Allowing for this possibility complicates the analysis substantially relative to Jovanovic and Moffitt (1990).

Kennan and Walker (2006) formulate a partial equilibrium model of optimal migration across U.S. states. In their model, as in my model and in Jovanovic and Moffitt (1990), workers need to migrate to a location in order to observe the value of their idiosyncratic match with it.[3] They exploit the panel structure of the NLSY data to identify wage differences among workers due to location matching effects. One of Kennan and Walker's contributions is to study return migration, an important issue from which I abstract in this paper in order to study migration decisions in general equilibrium.

Ex-ante heterogeneity among workers in mobility rates might represent an alternative explanation to learning about match quality by ex-ante identical workers (Jovanovic, 1979) to account for why most new jobs end early (or in this setting, for why newly arrived workers in a location tend to migrate with higher probability than workers already living there). Existing empirical evidence (see e.g. Flinn, 1986 and Farber, 1994) supports both kinds of effect and thus the model allows for both.

In my model workers care about consumption of goods and consumption of an immobile factor, land. The presence of land in the utility function is necessary in order for the model to generate the observed dispersion in average labor earnings across locations. From

unemployment. Topel (1986) considers a variant of the Lucas-Prescott model that allows for heterogeneity in experience among workers. Differently from these contributions based on a competitive setting, Lkhagvasuren (2005) adopts a Mortensen-Pissarides style of model to explain the existence of persistent differentials in unemployment rates across U.S. states.

[3] Miller (1984) and Flinn (1986) are earlier contributions that build on Jovanovic (1979)'s matching model. Differently from these papers, I abstract for simplicity from multiperiod learning effects: a worker is assumed to learn about his match with a location immediately after arriving there.

this perspective, the model generalizes Roback (1982)'s classic model of the effect of local amenities on local land prices and wages to a dynamic setting in which local amenities evolve stochastically over time, giving rise to the dynamics of local employment.[4] A related paper by Van Nieuwerburgh and Weill (2006) uses a version of Lucas and Prescott's island model to study the effect of increased wage dispersion across U.S. metropolitan areas on local housing prices. Differently from the latter, my paper focuses mainly on the dynamics of worker flows rather than on the behavior of average housing prices over time.[5]

On the empirical front, the paper also builds on the seminal contribution of Blanchard and Katz (1992), who document the existence of very persistent differences in employment growth rates across U.S. states. Differently from Blanchard and Katz who only consider net worker flows, this paper also focuses on gross flows.[6]

2 Data and Stylized Facts

In this section I briefly describe the data used in the paper and then organize the main features of worker flows into a series of stylized facts.

2.1 The Data

I use the Integrated Public Use Microdata Series (IPUMS) from the U.S. Census of Population for 1970, 1980, 1990, and 2000 (Ruggles et al., 2004).[7] The Census questionnaire includes a question regarding the state where an individual was living five years before the Census interview. Using this information, I construct rates of gross and net flows of population across the 48 contiguous United States.[8] The population flows always refer to the five year period preceding the Census year, and represent a lower bound on the actual flows, as some individuals moved more than once during these five years. In order to focus on geographic mobility that is not motivated by college attendance or retirement, I restrict attention to individuals who were between 27 and 60 years of age and in the labor force at the time of the Census. The sample includes both U.S. born workers as well as foreign-born ones who immigrated to the U.S. at least five years prior to the Census year. This restriction

[4] Roback (1982) allows for amenities to affect both the local production and utility functions. While in my benchmark model employment reallocation is driven by shocks to the production function, Appendix C shows how the model can be modified to allow for amenities that enter into workers' production functions.

[5] In a related paper, Morris and Ortalo-Magne (2007) study the joint distribution of housing prices and wages across metropolitan areas.

[6] The paper is also related to the research on the determinants of population flows within the U.S., surveyed by Greenwood (1975) and more recently developed by Greenwood and Hunt (1984), Treyz et al. (1993), Rappaport (2004), and Armenter and Ortega (2007).

[7] This is available online at `http://usa.ipums.org/usa/`. Appendix A contains more detailed information on the data, issues of sample selection, and on the construction of the variables used in the paper.

[8] The levels of inflow and outflow of population for a given state were standardized by the number of workers satisfying the sample selection criteria who were surveyed in the Census year and reported living in that state 5 years before. Net flow rates were defined as the difference between gross inflow and outflow rates.

guarantees that aggregate net flows of workers equal zero.[9] From now on, for simplicity, I will refer to a state's "population" as the collection of individuals satisfying these sample selection criteria.

Before proceeding, it is necessary to briefly comment on the choice of U.S. states as primary units of analysis. Since the focus of the paper is the *geographic* mobility of workers, the ideal unit of analysis should be a local labor market. The latter concept is intuitive but not simple to define unambiguously. In practice, a local labor market is often associated with a metropolitan area. In this paper I have chosen not to take a metropolitan area as the basic unit of analysis for several reasons. First, the 1970 Census does not report information on an individual's metropolitan area of residence in 1965. This information is instead available at the state level. This is important because the information contained in the 1970-2000 Censuses is used below to estimate the stochastic process for local labor demand shocks. The lack of the 1970 data would further reduce the time-series dimension of the data. Second, about 20 percent of the U.S. population did not live in a metropolitan area at the time of the 2000 Census. This figure has increased by about 10 percentage points since 1970, and it displays a non-trivial geographic variation. Third, a non-negligible number of metropolitan areas in the 1980-2000 Censuses were only incompletely identified, meaning that a subset of the households of a given metro area were not coded as living in that area. This creates problems as this subset of households is not random.

In summary, using a state as a unit of analysis, while in principle is less satisfactory than using a metropolitan area, allows for a more consistent definition of a "location" over time. This is important because the time-series dimension of the data provides useful information on the mechanism emphasized in the paper. Last, as mentioned in the introduction, the key fact on which I focus - the positive cross-sectional correlation between inflow and outflow rates - also holds at the level of metropolitan areas.

2.2 Stylized Facts

In what follows I introduce the main features of the data on worker flows using simple descriptive statistics.

2.2.1 Magnitude of Gross Flows

I start by summarizing the magnitude of gross and net flows. The statistics in the following tables are computed weighting each state by its relative population in order to avoid assigning disproportional importance to small outlier states. The first column of Table 1 reports gross worker reallocation - defined as the sum of gross inflows and outflow rates - across U.S. states. Column 2 represents the average absolute value of net flow rates across U.S. states, while column 3 reports the excess reallocation of workers, defined as the difference between columns 1 and 2.[10] Notice that gross flows are large relative to net flows. On average, a state gains or loses about 2.39 percent of its workforce in a 5-year period due to migration.

[9] In Appendix A.2, I show that recent immigration from outside of the U.S. does not affect significantly the stylized facts presented in Section 2.2.

[10] Appendix A provides further detail on the construction of these measures.

In the same period, the average state experiences a combined inflow and outflow of workers of 16.8 percent of its workforce.

Table 1
Magnitude of Worker Flows Across U.S. States

Census Year	Gross Worker Reallocation	Average Absolute Net Flow Rates	Excess Worker Reallocation
1970	13.86	1.74	12.12
1980	17.98	2.59	15.14
1990	18.12	2.87	15.25
2000	17.26	2.11	15.15
Average	16.80	2.39	14.41

2.2.2 Cross Sectional Correlations Among Worker Flows

The cross-sectional correlations among workers flow rates are reported in Table 2. The table shows partial correlation coefficients among worker flows computed using all state-year observations and after controlling for Census year dummies.

Table 2
Correlation Coefficients Among Worker Flows

	Outflow Rate	Inflow Rate
Net Flow Rate	0.02	0.84**
Inflow Rate	0.56**	

Note: ** significant at 1% level. * significant at 5% level.

The basic relationships between outflows, inflows, and net flows are evident from the scatter plots in Figures 1-3, which were constructed by pooling all state-year observations. Three features of the data stand out. First, as shown in the introduction, gross outflow rates tend to be relatively high in states characterized by relatively high gross inflow rates (Figure 1). Second, there is a U-shaped relationship between gross outflow and net flow rates (Figure 2), with the resulting correlation between these two variables being close to zero. It follows that states that tend to lose workforce due to migration do so by experiencing lower than average inflows, rather than higher than average outflows. For example, in the 2000 Census North Dakota and Nevada have both very large outflow rates (around 13-14 percent), even if they are at the opposite extremes of the distribution of net flow rates. Last, and less surprisingly, gross and net inflow rates are highly positively correlated (Figure 3).

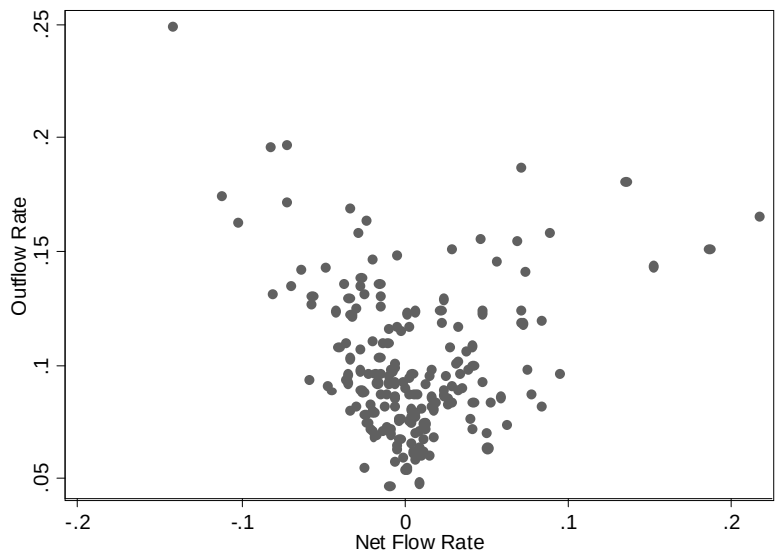

Figure 2: Scatter plot of net flow and outflow rates of workers across U.S. states, 1970-2000. The correlation coefficient between net flow and outflow rates is 0.02 and has been computed using year fixed-effects. Data source: my computations based on data from the U.S. Census of Population and Housing.

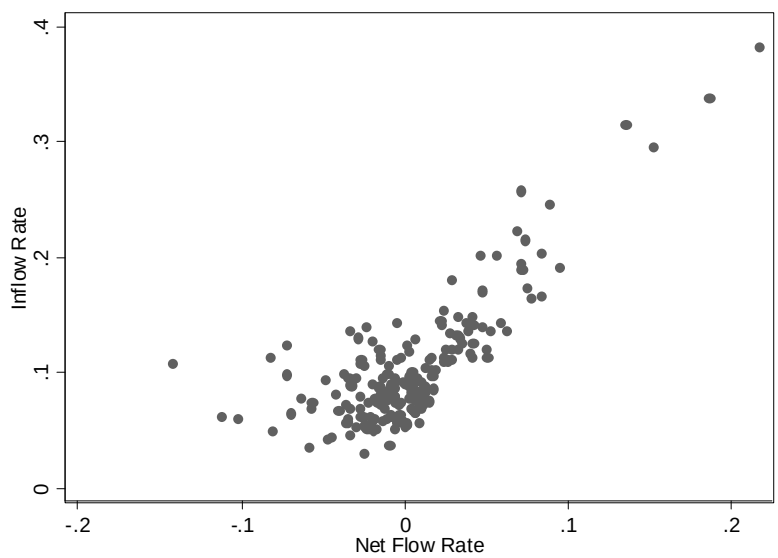

Figure 3: Scatter plot of net flow and inflow rates of workers across U.S. states, 1970-2000. The correlation coefficient between net flow and gross inflow rates is 0.84 and has been computed using year fixed-effects. Data source: my computations based on data from the U.S. Census of Population and Housing.

The positive cross-sectional correlation between gross inflow and gross outflow rates reported in Table 2 might be symptomatic of a changing industry or demographic mix of states' workforce (Sjaastad, 1962). To answer the question of whether gross worker flows occur mainly within or between demographic and industry groups, I divided the 2000 Census sample into 385 demographic groups defined according to age, education, and industry.[11] I then computed gross and net flow rates for each state and for each demographic group. A way to consider exclusively within-group flows is to calculate the cross-sectional correlations among worker flow rates separately for each of the 385 demographic groups. Table 3 reports the median of these correlation coefficients for the 2000 Census:[12]

Table 3
Adjusted Correlation Coefficients Among Worker Flows

	Outflow Rate	Inflow Rate
Net Flow Rate	-0.20	0.80^{**}
Inflow Rate	0.43^{**}	

Note: ** significant at 1% level. * significant at 5% level.

Notice that the correlations in Table 3 are almost the same as the ones in Table 2. In particular, the within-group correlation between inflow and outflow rates, while slightly lower than the value of 0.56 reported in Table 2, is still positive and statistically significant. Thus, for a given state, incoming workers tend to have a similar age and level of education and to work in the same industry as outgoing ones. This result is consistent with Miller (1967) who finds that inflow and outflow rates across U.S. metropolitan areas are positively correlated even after controlling for workers' sex, race, and occupation.

A complementary exercise which confirms these results is to compute for each U.S. state the fraction of excess worker reallocation that is due to between-group employment shifts. The population-weighted average of this measure across U.S. states for the 2000 Census was 0.08, suggesting that most flows occur within the demographic/industry groups described above. Similar results obtained for the other Census years.

Last, to further assess the nature of worker flows, it is possible to compare weekly earnings of workers who are migrating into a given state with the weekly earnings of workers who are leaving that same state. If outgoing and incoming workers belong to the same demographic groups we should not observe sizable differences between those. Without controlling for any observable worker characteristic, the average weekly labor earnings of a worker moving out of a state are about three percent lower than the weekly earnings of a worker moving into that same state. This difference is quite small, further confirming the within-group nature of gross migration flows.

Differences among observationally equivalent workers could be ex-post or ex-ante in nature. Ex-post heterogeneity might reflect idiosyncratic matching effects between workers and locations, as in Kennan and Walker (2006). Given that in Table 3 I am controlling for workers' age and education, ex-ante heterogeneity is likely to pertain to unobserved individual

[11] See Appendix A.1.3 for a detailed explanation of how these groups were constructed.

[12] Each of the correlations reported in Table 3 is the weighted median among 385 correlation coefficients. The weights used are the shares of the demographic groups' populations in the total population of migrants for the U.S. The significance level refers to the median correlation in Table 3.

attributes such as, for example, disutility of migration. The theoretical model of Section 3 incorporates both kinds of heterogeneity.

2.2.3 Worker Flows Over Time

Thus far I have focused on the cross-sectional correlations among worker flows. The advantage of using several Census years is that it is possible to assess the persistence of worker flows over time. Table 4 reports, for each type of flow, its first order autocorrelation coefficient across Census years, computed by pooling all state-year data points together.

Table 4
First-Order Autocorrelation of Worker Flows

Net Flow Rate	Outflow Rate	Inflow Rate
0.58**	0.69**	0.85**

Note: ** significant at 1% level. * significant at 5% level.

The autocorrelation coefficients imply that worker flows are very persistent across decades, suggesting that low-frequency shocks play a major role in reallocating employment across U.S. states. This finding is in accordance with the evidence on states' employment growth presented by Blanchard and Katz (1992). More subtly, notice that gross inflow rates are more highly autocorrelated over time than net flow rates. In Section 5 below, I show that this difference is consistent with the selection effects highlighted in the paper.

3 Model

In order to account for these stylized facts I now introduce a general equilibrium model of gross and net migration flows.

3.1 Description

The model builds on the island-model of the labor market developed by Lucas and Prescott (1974) and on Roback (1982)'s static analysis of workers' and firms' location decisions. The force that drives the dynamics of the local labor and land markets in the model is a persistent shock to total factor productivity. The latter generates temporary increases in local wages and land prices that are then followed by net inflows of workers. Simultaneously, idiosyncratic match shocks give rise to workers' gross flows. In equilibrium, the value of migrating from one labor market to another is pinned down by the requirement that aggregate net flows of workers are zero.

Production and Firms. The economy is populated by a continuum of locations ("islands") of measure one, indexed by $j \in [0, 1]$. Each location is endowed with one unit of a local and immobile factor of production ("land"). There is only one good in this economy whose price is normalized to unity. The good is produced in each location by a large number of competitive firms, each endowed with the constant returns to scale production technology

$zF(y, l^f)$. The production input y represents the units of labor located in the island while l^f is land used in production.[13] The production function F is concave and is characterized by the following Inada conditions: $\lim_{y\to 0} F_y(y, l^f) \to \infty$ and $\lim_{l^f \to 0} F_l(y, l^f) \to \infty$. The variable z represents total factor productivity in the island. Firms solve static optimization problems hiring labor at the rate w and units of land at the rate r after observing the productivity shock z.

Exogenous Shock. The law of motion for z' depends on z and another random variable ε. Letting $\zeta = (z, \varepsilon)$, the exogenous state vector for a location is assumed to evolve over time according to a stationary Markov process with transition function $Q(\zeta', \zeta)$. The exact details of this process will be specified in Section 4.

There are two complementary ways of thinking about the nature of the shock z. According to the first, z captures location-specific productivity differentials that in principle affect different industries in the same way. Examples of such factors are state-level corporate and union legislations. Holmes (1998) provides empirical evidence supporting the view that right-to-work laws and other probusiness policies (such as weak environmental and safety regulations) have a positive causal effect on the concentration of manufacturing activity across U.S. states.[14] Clark (1998) finds evidence of "significant region-specific components in the cyclical variation" of employment growth fluctuations across U.S. Census regions after controlling for industry mix effects.

According to a second view, z is not a productivity shock to all sectors in a location, but instead represents, in a reduced-form way, the effect that sectoral shocks have on the local economy, once local intersectoral interactions are accounted for. For example, the decline of the steel industry in Pennsylvania was also reflected in lower demand for local services (education, health, etc.) and construction. While the original shock affected one industry, over a sufficiently long time-horizon the disaggregated data showed a broad decline in employment across most sectors of the local economy.[15]

While the model assumes that productivity shocks are the main driving force behind employment reallocation across U.S. states, it is possible to modify it to allow for stochastic "amenity" shocks that affect workers' utility in a location (e.g., the introduction and diffusion of air conditioning made hot and humid summers in the south-western part of the U.S. more tolerable). Appendix C shows that this version of the model would have exactly the same implications for migration flows across locations as the model considered in this section. The

[13] The production function could also be made a function of physical capital. Under the assumption of perfect capital mobility, the latter would move to equalize its rate of return across locations. In this case, the function F in the text should be interpreted as the reduced form production function obtained by solving out for the optimal amount of capital in a location and replacing the latter back into the original production function. I take physical capital into account when calibrating the model in Section 4. See Rappaport (2004) for the analysis of a two-location model economy with adjustment costs to physical capital.

[14] More generally, in the post-WWII period the manufacturing industry has progressively moved away from the North-East and toward the South-West of the U.S., attracted by more favorable union and corporate legislation and fiscal incentives (see Peet, 1983, Cobb, 1993, and English, 2006 for an historical perspective on this issue).

[15] For example, using yearly employment data on one-digit industry employment by U.S. state from the Bureau of Labor Statistics, 1969-2001, I find that *national* trends in employment shares by industry can explain about 40 percent of the variance of employment growth across states over a one year period. However, over a five year period this figure drops to 7 percent, and over a ten year period to only 1 percent.

model's focus on labor demand rather than labor supply shocks as a source of employment reallocation across locations is consistent with the positive cross-sectional correlation between state-level earnings and rents. The cross-sectional correlation between these two variables, computed by pooling all state-year data together and controlling for year dummies, is 0.80 and highly significant.[16]

Ex-Ante Heterogeneity. The economy is populated by a measure one of infinitely-lived dynasties. At a point in time only one worker is alive in each dynasty. In order to allow for ex-ante heterogeneity in mobility rates across workers, I assume that there are two types of dynasties. To keep the model simple, I assume that a measure θ of the dynasties are comprised by workers who always exogenously relocate in every period. A measure $1-\theta$ of dynasties, instead, includes workers who do not relocate exogenously, but instead choose to migrate or not based on idiosyncratic shocks and the state of the local economy. In what follows, I describe the preferences, constraints, and the decision problem of this latter type of dynasty.

Preferences, Timing, and Ex-Post Heterogeneity. A worker who belongs to the dynasty in which there is no exogenous relocation has the following instantaneous utility function:[17]

$$u = c + \phi\left(l^c\right) - \upsilon, \tag{1}$$

where c denotes consumption of goods, l^c represents consumption of land, and υ is an idiosyncratic shock to utility that summarizes a worker's match with the location in which she lives.[18] The function ϕ is such that $\phi' > 0$, $\phi'' < 0$, and $\lim_{x\to 0}\phi'(x) \to \infty$. An agent i's flow budget constraint in period t and location j is:[19]

$$w_{jt} = c_t + r_{jt} l_t^c.$$

Workers discount future utility at the rate $\beta < 1$.

In detail, the sequence of events in the life of a worker whose probability of exogenous relocation is zero (recall that these comprise a fraction $1-\theta$ of the population) is as follows:

- An agent is born in a location at the end of period $t-1$.
- At the beginning of t, the agent draws the idiosyncratic match shock υ. For simplicity, the shock υ is assumed to take only two values:

$$\upsilon = \begin{cases} \upsilon_1 \text{ w.p. } p \\ \upsilon_2 \text{ w.p. } 1-p \end{cases}. \tag{2}$$

[16] State-level earnings and rents are measured as state fixed-effects in hedonic regressions of weekly earnings and rents on observable characteristics of workers and renter-occupied housing units. See Appendix A for details.

[17] Van Nieuwerburgh and Weill (2006) consider a similar specification for the static utility of the agent as a function of consumption of goods and housing services. In their model, though, housing is elastically supplied and does not enter into the firms' production function.

[18] In a previous version of the paper, I adopted Kennan and Walker (2006) specification and assumed that workers first moved into a location and then discovered their efficiency units of labor there. These two versions of the model produce nearly identical quantitative implications.

[19] Notice that this budget constraint does not include land income. This is without loss of generality given the specification of the instantaneous utility function adopted here.

The idiosyncratic shock υ represents an agent's match with the location and remains the same as long as the agent stays in the same location.

- The agent optimally chooses c and l^c and receives a utility flow given by equation (1).
- With probability $1-\delta$ the agent dies and is replaced by another agent of the same dynasty that starts her life in a random location at the end of period t. In this random assignment, the probability that a newly born agent starts her economic life in a given location at the end of period t is assumed to be proportional to that location's population at the beginning of period t. With probability $\delta < 1$ the agent survives into the next period.
- If the worker survives the death shock, she can decide whether to stay in the same location or move to another location. The information available to the agent when making this choice will be specified later. If she decides to move she obtains expected utility e.
- At the beginning of period $t+1$, if the agent had remained in the same location in which she was living in t, she keeps the same idiosyncratic shock υ. If the agent had chosen to move to a new location, she pays a utility cost k, and draws a new idiosyncratic location-match from the distribution in equation (2).

Search. The literature has typically made two different kinds of assumptions about the nature of search in this class of models. One approach is to make search directed, so that a migrating agent relocates to the location that offers the highest expected utility. An alternative (see e.g. Kambourov and Manovskii, 2004) is to make search undirected and therefore assume that agents are randomly reallocated across locations.[20] Here, I allow for both possibilities. Specifically, with probability η a migrating agent is directed toward the location that offers the highest expected utility, denoted by e_d. The timing of the model is such that an agent must decide whether to migrate or not from a certain location before the realization of next period's aggregate shock z' in that location. An agent that has decided to migrate, and has to determine where to direct herself, is assumed to know only the expected realization of the shock z' in all potential locations of choice. With probability $1-\eta$, instead, the agent is randomly reallocated and obtains expected utility e_r. Also in this case, as for newly born individuals, the probability of arriving to a given location at the end of period t is assumed to be proportional to that location's population at the beginning of period t. By definition, then:

$$e \equiv \eta e_d + (1-\eta) e_r.$$

Intuitively, if $\eta = 0$ (undirected search), inflow rates will be the same for all locations. Variations in outflow rates will accommodate changes in net flow rates, and these two variables will tend to be strongly negatively correlated in the cross-section. At the other extreme, if $\eta = 1$ (directed search), the model tends to generate a positive and large cross-sectional correlation between net flow and outflow rates. A continuity argument, therefore, implies

[20] A natural interpretation of "undirected search" in models of geographical mobility is that an agent is drawn to a given location by idiosyncratic factors, such as the presence of family and friends.

that for middle values of η the model can reproduce the observed correlation between gross outflows and net flows, which is about zero in the data (see Table 2). In Section 4 below I estimate the parameter η in order to reproduce the observed value of this correlation.

Value Function. At the beginning of a period the state of a location is fully characterized by the vector (y, n, m, ζ), where n denotes the measure of workers with match value υ_1, and m denotes the measure of workers who always migrate for exogenous reasons. Denote the laws of motion for y, n, and m by $Y(s,\zeta)$, $N(s,\zeta)$, and $M(s,\zeta)$ respectively, and summarize the three in the law of motion for $s = (y, n, m) : s' = S(s,\zeta)$, where $S(s,\zeta) = [Y(s,\zeta), N(s,\zeta), M(s,\zeta)]$. The dynamic programming problem of a worker characterized by idiosyncratic match υ with the location is given by:

$$\begin{aligned} V(s,\zeta,\upsilon) &= \max_{c,l^c} \left\{ c + \phi(l^c) - \upsilon + \beta\delta \max\left[\int V(s',\zeta',\upsilon)\, Q(\zeta, d\zeta'), e - k \right] + \beta(1-\delta)\, e_r \right\} \\ & \quad s.t. \qquad (3) \\ w(s,\zeta) &= c + r(s,\zeta)\, l^c, \\ s' &= S(s,\zeta). \end{aligned}$$

Inflows and Outflows. Denote gross inflows into a location characterized by state (s,ζ) by $x(s,\zeta)$ and gross outflows from that location by $o(s,\zeta)$. Gross inflows can be written as:

$$x(s,\zeta) = x_d(s,\zeta) + \overline{x}(1-\eta)\, y, \tag{4}$$

where the first term on the right-hand side of this equation represents the component of gross inflows due to directed search while the second term represents the one due to undirected search. The latter is equal to the product of the aggregate level of inflows $\overline{x}(1-\eta)$ that are undirected and the share y of those assumed to flow into the location.

Let $\delta n q(s,\zeta,\upsilon_1)$ denote the measure of workers with idiosyncratic shock υ_1 that chooses to leave the location and by $\delta(y-n-m)\, q(s,\zeta,\upsilon_2)$ the equivalent measure of workers with idiosyncratic shock υ_2. Outflows $o(y,\zeta)$ are then equal to:

$$o(s,\zeta) = \delta n q(s,\zeta,\upsilon_1) + \delta(y-n-m)\, q(s,\zeta,\upsilon_2) + \delta m, \tag{5}$$

where the last term on the right-hand side of this equation represents out-migration by the type of workers who are assumed to always migrate in each period.

Then, the laws of motion $Y(s,\zeta)$, $N(s,\zeta)$, and $M(s,\zeta)$ can be written as:

$$\begin{aligned} Y(s,\zeta) &= y + x(s,\zeta) - o(s,\zeta), && (6) \\ N(s,\zeta) &= \delta n(1 - q(s,\zeta,\upsilon_1)) + p(1-\delta\theta/\overline{x})\, x(s,\zeta) + p(1-\theta)(1-\delta) y, && (7) \\ M(s,\zeta) &= (\delta\theta/\overline{x})\, x(s,\zeta) + y\theta(1-\delta). && (8) \end{aligned}$$

Equation (6) states that changes in the workforce equal the difference between inflows and outflows.[21] Equation (7) gives the measure of workers characterized by idiosyncratic

[21] Notice that by construction there is no change in the workforce due to the death/birth process: workers who die are replaced by newly born ones who are distributed across locations proportionately to their initial relative size.

shock υ_1 at the beginning of next period as the sum of three terms. The first term on the right-hand side of (7) represents the measure of workers who drew the shock υ_1 in the previous period and decided not to migrate. The second term represents the share p of workers who migrates into the location and draws the shock υ_1. Notice that while $x(s,\zeta)$ represents the total measure of workers migrating in the location, only a fraction $(1-\delta\theta/\overline{x})$ of those workers belong to dynasties that are not exogenously relocated in each period. The third term on the right-hand side of equation (7) represents the fraction p of newly born workers who start their lives in the location. Only a fraction $1-\theta$ of those belong to dynasties that are not exogenously relocated in each period. Equation (8) states that at the beginning of next period the measure of workers who always relocate exogenously equals to the fraction $(\delta\theta/\overline{x})$ of inflows $x(s,\zeta)$ plus a fraction θ of newly born workers who start their lives in the location.

Stationary Distribution. I consider a stationary environment with a location-invariant distribution of workers across locations $\mu(s,\zeta)$. The latter is such that:

$$\mu(S',\Xi') = \int_{\{s,\zeta:s'\in S'\}} Q(\zeta,\Xi')\mu(ds,d\zeta). \tag{9}$$

3.2 Equilibrium

A recursive stationary equilibrium for this economy is represented by a value function $V(s,\zeta,\upsilon)$, a probability $q(s,\zeta,\upsilon)$, laws of motion $S(s,\zeta) = [Y(s,\zeta), N(s,\zeta), M(s,\zeta)]$, gross inflows $x(s,\zeta)$, gross outflows $o(s,\zeta)$, the expected utility of being randomly reallocated e_r, the expected utility of directed migration e_d, an aggregate level of inflows $\overline{x}$, a stationary distribution $\mu(s,\zeta)$, demand for land by consumers $l^c(s,\zeta)$, demand for land by firms $l^f(s,\zeta)$, a wage function $w(s,\zeta)$, and a rent function $r(s,\zeta)$ such that:

- The value function $V(s,\zeta,\upsilon)$ satisfies the Bellman equation (3) given e_r, e_d and the law of motion $S(s,\zeta)$. In addition, the demand for land by consumers in location (s,ζ) satisfies the first order condition:

$$r(s,\zeta) = \phi'(l^c(s,\zeta)). \tag{10}$$

- The law of motion $S(s,\zeta)$ is related to $x(s,\zeta)$, $o(s,\zeta)$ and $q(s,\zeta,\upsilon)$ by equations (6)-(8).

- Inflows $x(s,\zeta)$ are consistent with directed search by migrating workers:

  - If $x(s,\zeta) = y\overline{x}(1-\eta)$ then:

$$e_d \geq p\int V(S(s,\zeta),\zeta',\upsilon_1)Q(\zeta,d\zeta') + (1-p)\int V(S(s,\zeta),\zeta',\upsilon_2)Q(\zeta,d\zeta').$$

  - If $x(s,\zeta) > y\overline{x}(1-\eta)$ then:

$$e_d = p\int V(S(s,\zeta),\zeta',\upsilon_1)Q(\zeta,d\zeta') + (1-p)\int V(S(s,\zeta),\zeta',\upsilon_2)Q(\zeta,d\zeta').$$

- Outflow probabilities $q(s,\zeta,\upsilon)$ are consistent with individual optimization:
  - If $q(s,\zeta,\upsilon)=0$ then:
$$\int V\left(S\left(s,\zeta\right),\zeta',\upsilon\right)Q(\zeta,d\zeta')\geq\eta e_d+(1-\eta)e_r-k,$$
  - If $q(s,\zeta,\upsilon)=1$ then:
$$\int V\left(S\left(s,\zeta\right),\zeta',\upsilon\right)Q(\zeta,d\zeta')\leq\eta e_d+(1-\eta)e_r-k,$$
  - If $q(s,\zeta,\upsilon)\in(0,1)$ then:
$$\int V\left(S\left(s,\zeta\right),\zeta',\upsilon\right)Q(\zeta,d\zeta')=\eta e_d+(1-\eta)e_r-k.$$
- The value of being randomly reallocated e_r is:
$$e_r=p\int V\left(s,\zeta,\upsilon_1\right)y\mu\left(ds,d\zeta\right)+(1-p)\int V\left(s,\zeta,\upsilon_2\right)y\mu\left(ds,d\zeta\right).$$
- Aggregate population has measure one:
$$\int y\mu\left(ds,d\zeta\right)=1.$$
- The invariant distribution $\mu\left(s,\zeta\right)$ is consistent with individual decisions, so equation (9) holds.
- The wage and rent functions in each location are such that:
$$\begin{aligned} r(s,\zeta) &= zF_l\left(y,l^f(s,\zeta)\right), \\ w(s,\zeta) &= zF_y\left(y,l^f(s,\zeta)\right). \end{aligned} \tag{11}$$
- The land market in each location clears:
$$yl^c(s,\zeta)+l^f(s,\zeta)=1. \tag{12}$$
- Aggregate gross inflows are given by:
$$\overline{x}=\int x(s,\zeta)\mu\left(ds,d\zeta\right).$$

4 Empirical Implementation

When bringing the model to the data, one has to keep in mind that the model assumes the existence of a continuum of locations, and therefore a constant expected utility of migration e. The assumption of a continuum of location is mainly for feasibility: allowing for a finite number of locations in the theoretical model would make it virtually impossible to solve because a worker in a location would have to take into account the distribution and dynamics of the state vector (s, ζ) across all other locations when solving her dynamic programming problem.

Some of the model's parameters are set a-priori and others are estimated using a version of the method of simulated moments (see Lee and Ingram, 1991 and Duffie and Singleton, 1993). Consider first the parameters that are set a-priori. A period in the model is taken to represent 5 years. The discount factor β is set equal to 0.82, implying a yearly interest rate of 4 percent. The parameter δ determines the size of the cohort entering the labor market in each period. In the data workers in the age group 27-32 represent about 15 percent of the sample, which implies $\delta = 0.85$.

The parameters that determine ex-ante and ex-post heterogeneity in the population are set to strike a balance between empirical plausibility and computational considerations. The random variable υ can take one of two values, υ_1 and υ_2, with probability p and $1-p$, respectively. The value of the shock υ_2 is normalized to zero. I pick a value for υ_1 that is large enough so that poorly matched workers always migrate independently of the state of the local economy where they reside (i.e., $q(s, \zeta, \upsilon_1) = 1$). Since there are many values of υ_1 that are consistent with this condition, I select the approximate minimum from this set, $\upsilon_1 = 0.08$, by solving analytically the steady state version of the model, and then checking ex-post that agents drawing υ_1 always choose to migrate in each possible state of the local economy, once location-specific shocks are introduced.[22]

The parameter p is not set-ex-ante but rather estimated in order to match the average migration rate across all Census years (see below). Census data provide little guidance on how to set the share θ of dynasties who always migrate. This is because the Census is a cross-section and it only provides information about one move per worker. Kennan and Walker (2006) analyze longitudinal data on young workers and report that in their sample about 82 percent of workers never move in a period of 13 years, or about 3 model periods. I use this information to compute an approximate value of θ as a function of p in the following way. Consider a just-born worker of a dynasty that is not always exogenously reallocated. With probability $1-p$ this worker draws a "good" match with the location and does not choose to migrate for idiosyncratic reasons. Thus, the unconditional probability she *never* migrates for idiosyncratic reasons is $(1-\theta)(1-p)$. However, she might migrate for location-specific reasons even if her idiosyncratic match with the location is good. From Table 1, one obtains that the migration rate for location-specific reasons is about 1.2 percent over a 5-year period.[23] Thus, setting the probability that a worker never moves in 3 model periods equal

[22] Appendix B provides details on the characterization of the model's equilibrium in this situation in which poorly matched workers always choose to migrate.

[23] This figure is obtained by dividing the average absolute net flow rate (2.39 percent in Table 1) by two.

to the Kennan-Walker figure yields the following implicit relationship between θ and p:

$$(1-\theta)(1-p)(1-0.012)^3 = 0.82.$$

This equation determines the parameter θ as a function of p.

The structure of the model and the cross-sectional nature of the Census data do not allow me to identify the magnitude of the moving cost parameter k. Instead, the latter is set equal to 0.1. This represents 11.5 percent of the average wage income in the model. To place this value in perspective, notice that the average labor income earned by a worker in the 2000 Census sample I consider is about 44,000 in 2005 US$. Given that one period in the model represents 5 years, this implies that the moving cost is about 25,000 expressed in 2005 US$.[24]

The production technology and the utility function are assumed to take the following forms:

$$F(y,l) = y^{\tau} l^{1-\tau},\ \tau \in (0,1), \tag{13}$$

$$\phi(l) = \frac{A}{\alpha} l^{\alpha},\ \alpha \in (-\infty, 1),\ A > 0. \tag{14}$$

The parameter τ represents the share of labor income in total output. To calibrate this parameter, it is necessary to take into account the fact that physical capital has already been solved out of the profit optimization problem of the firm. The parameter τ is set equal to 0.9091. I obtain this number using Caselli and Coleman (2001)'s computation of the share of land and labor in the manufacturing sector. The elasticity parameter α is set equal to $1-\tau$. This value greatly facilitates the numerical solution of the model because it implies that land can be solved out of the model analytically (see Appendix B).

The transition function $Q(\zeta', \zeta)$ is assumed to take the form:

$$z' = z\varepsilon', \tag{15}$$

$$\varepsilon' = z^{\psi-1}\varepsilon^{\rho} u' \tag{16}$$

where $\psi < 1$, $\rho < 1$, and u' is independent and identically distributed both over time and across locations according to a lognormal distribution with mean one and variance $(\exp\{\sigma_u^2\} - 1)$. The specification of the exogenous shocks in equations (15)-(16) is somewhat non-standard because it assumes that the growth rate of productivity ε' is persistent. This is necessary in order to generate persistent net flows in the Lucas-Prescott model. If ρ were equal to zero, net flows would be negatively autocorrelated over time, which is strongly at

[24] Kennan and Walker (2006) obtain estimates of moving costs between \$229,000 and \$176,000 in 2005 \$. Calibrating the model to a higher moving cost while simultaneously adjusting the value of υ_1 to guarantee that poorly matched workers always choose to migrate does not substantially alter the quantitative results of the paper. The only major difference is that this version of the model tends to yield a positive, rather than zero, cross-sectional correlation between net flow and outflow rates. This is because workers with idiosyncratic shock υ_2 never choose to migrate if moving costs are so large.

odds with the data (see Table 4).[25] The parameter ρ is estimated (see below). In order to obtain a stationary process for $\{z\}$ it is necessary that the parameter ψ be strictly smaller than one. The value of the parameter ψ is set exogenously equal to 0.999. This value is such that the growth rate of the shock process z is approximately an AR(1), while at the same time preserving the stationarity of z.

The remaining vector of parameters to be estimated is then $\phi = (\rho, \sigma_u, p, A, \eta)$. First, ρ and σ_u are estimated by matching some of the cross-sectional and time-series moments of states' net inflow rates reported in Section 2.2. Specifically, the parameter σ_u is identified by the average (across U.S. states and the four Census decades) absolute value of net inflow rates, whose value is 2.39 percent (Table 1). The parameter ρ is identified by the first-order autocorrelation coefficient of net inflow rates across Census years. The value of this coefficient is 0.58 (Table 4).

Second, the parameter p determines the probability of drawing a low idiosyncratic shock. Since agents drawing these shocks always choose to migrate, the parameter p is set to match the average (across the four Census years) interstate migration rate in the U.S. economy. From Table 1, this value is 8.40 percent.

Third, the parameter A is set to match the standard deviation of the logarithm of state-level weekly labor earnings across states and Census years. State-level labor earnings for a given Census year are measured as state fixed-effects in cross-sectional regressions of log weekly earnings on workers' observable characteristics (see Appendix A for details). According to this measure, the dispersion in labor earning across states is 9.29 percent. In the model, a higher value of A implies that the fixed factor ("land") becomes relatively more important in workers' utility. In equilibrium, locations with higher productivity z are characterized by higher rents because these locations are more attractive to firms. Since workers are also affected by the higher price of the fixed factor, these locations need to offer higher wages as well in order to attract the desired amount of labor. Thus, a higher value of A translates immediately into a higher dispersion of observed wages.

Last, the parameter η determines the extent to migrating workers choose to locate on the basis of the aggregate state of alternative locations (directed search), or on the basis of unmodelled idiosyncratic factors (undirected search). This parameter is identified by the cross-sectional correlation between net flow and outflow rates, which is equal to 0.02 (see Table 2). On the one hand, a higher value of η tends to increase the correlation between outflow and net flow rates as poorly matched workers and workers who always relocate tend to migrate out of locations that attract a disproportionate number of migrants. At the opposite extreme, when $\eta = 0$, and search in entirely undirected, all locations receive the same proportion of migrants. Thus, the only channel of adjustment of a location to shocks is represented by variations in outflow rates, which are relatively high in locations with negative net flows and low in locations with positive net flows. Thus, when η is close to zero, net flow

[25] An alternative way of obtaining persistent net flows would be to introduce less-than-perfect capital mobility in the model, as opposed to the current setting in which capital is assumed to be perfectly mobile. The parameter governing capital adjustment costs would then determine the extent of autocorrelation in net flows. Given the lack of independent evidence on the magnitude of these costs, I choose the simpler specification in which the shock process is characterized by persistent innovations.

and gross outflow rates are negatively correlated. The estimation procedure pins down the value of η that yields the observed correlation between outflow and net flow rates.

The following table summarizes the calibrated and estimated values of the model's parameters:

Table 5
The Model's Parameters

	Parameter	Estimate
β	discount factor	0.8200
δ	probability of death	0.8500
τ	production function parameter	0.9091
α	utility function parameter	0.0909
A	utility function parameter	0.0834
υ_1	value of low idiosyncratic shock	0.0800
υ_2	value of high idiosyncratic shock	0.0000
k	moving cost	0.1000
η	search parameter	0.3467
ψ	mean-reversion parameter	0.9990
ρ	persistence of labor demand shock	0.8360
σ_u	volatility of labor demand shock	0.0034
p	probability of low idiosyncratic shock	0.0962
θ	share of dynasties who always relocate	0.0593

Several features of the parameters' estimates stand out. First, the growth rate of productivity is found to exhibit very large persistence with a first-order autocorrelation coefficient of about 0.83 at 5-year intervals. Notice that if the productivity shock z, rather than its growth rate ε, had followed an AR(1) process, the first order autocorrelation coefficient of net flow rates would have been slightly negative, which is strongly at odds with the data.

Second, the estimate of the parameter A implies that a worker spends, on average, 9 percent of her labor income on the fixed factor. This is consistent with Roback (1982) who finds that the budget share of land in the data is about 6 percent.

Third, the estimate of η suggests that directed search accounts for about one third of all moves. To the best of my knowledge, there are no available estimates of this parameter in the labor or macro literatures that would allow for a comparison with the figure obtained here.

Fourth, the share of workers that always relocate accounts for about 5.93 percent of the population. Given an average migration rate of 8.4 percent, these workers account for about 70 percent of the moves in a 5-year period. Since migration due to local economic conditions accounts for about 14 percent of the moves, ex-post heterogeneity due to matching effects explains the remaining 16 percent.

5 Model's Fit

In this section I present the main quantitative results of the model. The first dimension along which I evaluate the model's fit is the relationship between net flow rates, on the one hand, and labor earnings and land rents, on the other. While the paper focuses on changes in the quantity of labor located in each state, it is important to check whether its predictions for prices are also consistent with the data. Given the lack of state-level price indices, I focus on nominal labor earnings and land rents, where the latter is meant to capture the price of the local immobile input used for both production and consumption purposes. Specifically, state-level labor earnings and land rents are measured as state fixed-effects in hedonic regressions of individual earnings and rents on observable characteristics of workers and renter-occupied housing units (see Appendix A for details). Table 6 presents the summary statistics on labor earnings, land rents and worker net flow rates.[26]

Table 6
State-Level Earnings, Rents, and Net Flow Rates

	Cross-Sectional Correlations				Spearman	
	Rents		Net Flow Rates		Rank Correlation	
	Data	Model	Data	Model	Data	Model
Earnings	0.80	0.96	−0.06	0.15	0.77	0.99
Rents			−0.06	0.02	0.76	0.99

Both in the model and in the data, the cross-sectional correlation between net flow rates and average earnings and rents is quite small. The intuition for this result is that net flow rates depend on the *growth rate* of total factor productivity in a location, ε, while wages and rents depend on the *level* of productivity, z. In locations with higher levels of productivity land rents and wages must be higher than in locations where productivity is low to make both firms and workers indifferent with respect to their choice of location (Roback, 1982). Thus, wages and rents are positively correlated in the cross-section. Moreover, the rank correlation coefficients between state-level earnings and rents in one Census year and the equivalent measures ten years before (in the previous Census year) are positive and large both in the data and in the model, suggesting considerable persistence in states' rankings. This parallels the persistence observed in worker flows.

The cross-sectional correlation patterns among worker flows in the benchmark model and in the data are reported in Table 7. Here and in what follows, I emphasize in bold the moments that were targeted in the estimation of the model.

Table 7
Correlation Coefficients Among Worker Flows

	Outflow Rate		Inflow Rate	
	Data	Model	Data	Model
Net Flow Rate	**0.02**	**0.02**	0.84	0.77
Inflow Rate	0.56	0.65		

[26] The cross-sectional correlations in the data were computed by pooling all state-year observations together and controlling for year dummies.

As the table shows, the model can account quite well for the positive correlation between gross inflow and outflow rates. Also, consistently with the empirical evidence the almost-zero correlation between gross outflow and net flow rates conceals a U-shaped relationship between these two variables. This is represented in Figure 4, which plots gross flow rates against net flow rates using data generated by the benchmark version of the model.

The intuition behind these results is as follows. As shown in Figure 4, locations characterized by positive net flows tend to display relatively high gross inflow rates and also relatively high gross outflow rates. The positive correlation between inflow and outflow rates in this region is due to the selection mentioned in the introduction, as incoming workers tend to have a higher propensity to migrate from the location to which they have arrived than workers already living there. Both ex-ante and ex-post heterogeneity among workers tend to produce this effect. First, the pool of in-migrating workers is disproportionately represented by workers characterized by ex-ante high mobility rates who will move again after spending one period in the location. Second, upon arrival incoming workers draw their idiosyncratic match with the location from the unconditional distribution of υ. It follows that a fraction p of these will draw a low idiosyncratic match with the location and choose to migrate again at the end of the period.

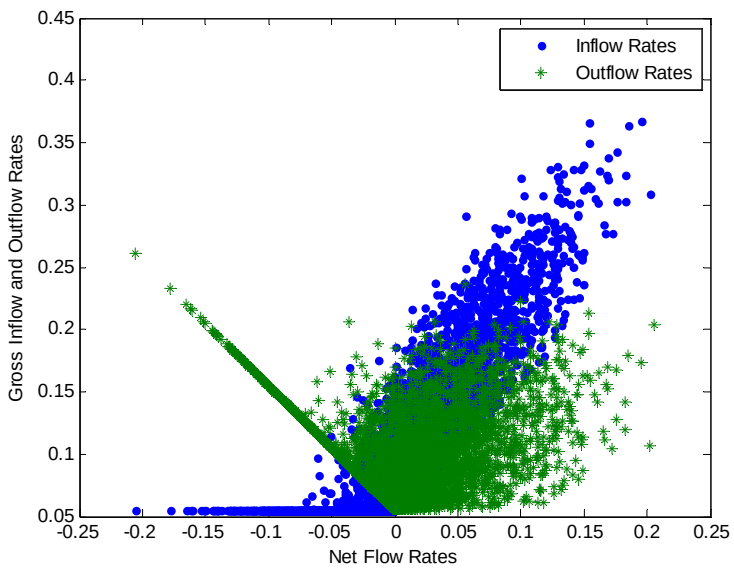

Figure 4: Scatter plot of gross inflow rates against net flow rates using data generated by the benchmark version of the model (10,000 observations).

To understand the decreasing side of the U-shaped relationship between net flow and outflow rates, notice that, as net flow rates become negative, gross inflow rates reach their lower bound (i.e., $x_d(s, \zeta) = 0$), and the location adjusts to further negative productivity shocks by means of higher gross outflows. In this region, also some of the workers who have a good idiosyncratic match with the location choose to migrate. Hence, gross outflow rates are relatively large for both large positive and large negative net flow rates.

Table 8 represents the first-order autocorrelation coefficients of net and gross worker flows.

Table 8
First-Order Autocorrelations of Worker Flows

Net Flow Rate		Outflow Rate		Inflow Rate	
Data	Model	Data	Model	Data	Model
0.58	**0.58**	0.69	0.62	0.85	0.80

The moments implied by the model are quite close to their empirical counterpart. Specifically, the model correctly predicts that gross inflow rates are more persistent than net flow rates. This is what one would expect if recent migrants into a location had a higher propensity to migrate than workers already living there. In this case relatively large inflow rates in the current period would give rise to relatively large outflow rates in the future. In turn, higher than average outflows would further lead to higher inflows in order to accommodate a given process of net flows. To see this formally, notice that by definition of inflow rate, the covariance between a location's inflows in t and in $t-1$ can be written as:

$$\begin{aligned}\text{covariance}\left(\text{inflow rate}_t, \text{inflow rate}_{t-1}\right) &= \text{covariance}\left(\text{outflow rate}_t, \text{inflow rate}_{t-1}\right) \\ &\quad + \text{covariance}\left(\text{net flow rate}_t, \text{inflow rate}_{t-1}\right).\end{aligned}$$

Now, assume for sake of illustration that the net flow rate is i.i.d. over time for a given location, in which case the second covariance term on the right-hand side of this equation is equal to zero. In this case, a positive correlation between current outflows and lagged inflows implies a positive autocorrelation of inflow rates, despite the fact that the underlying process for net flows displays no persistence.[27]

6 Additional Evidence on the Mechanism

A key prediction of the model is that states with large inflow rates in one period should experience relatively large outflow rates in *subsequent* periods, rather than contemporaneously. In other words, the observed positive contemporaneous correlation between gross inflow and outflow rates documented in Tables 2 and 3 should really be interpreted as the by-product of the positive correlation between current outflows and lagged inflows combined with the high persistence displayed by gross inflow rates (Table 4). I can evaluate the extent to which this argument is correct by using the time-series dimension of worker flow data across several Census decades. Ideally, one would want to regress the outflow rates between years $T-5$ and T on both the contemporaneous inflow rates between $T-5$ and T and the lagged inflow rates between $T-10$ and $T-5$. In order for the data to be consistent with the model, the current outflow rate should display a higher correlation with the *lagged* inflow rate than

[27] In the version of the model without heterogeneity among workers $(\theta = p = 0)$, and the same remaining parameters as the benchmark model, the first-order autocorrelation coefficient of gross inflow rates is basically the same as the one for net flow rates.

with the current inflow rate. Unfortunately, given that Census data are only available once every decade, the second-best option is to use, as proxy for lagged inflows, the inflows that occurred between year $T-15$ and year $T-10$, instead of the one between $T-10$ and $T-5$. The results of this exercise are reported in Table 9.

Table 9
Current Outflows and Lagged Inflows

	Gross Outflow Rate$_{T-5,T}$
Gross Inflow Rate$_{T-5,T}$	-0.04
Gross Inflow Rate$_{T-15,T-10}$	0.33^{**}
Year dummies	yes
Number of observations	192
R^2	0.37

Note: ** significant at 1% level. * significant at 5% level.

The results are quite clear. Controlling for lagged inflows, current inflows display no significant correlation with current outflows. This evidence is consistent with the mechanisms highlighted in this paper.

7 Ex-Ante vs Ex-Post Heterogeneity

Given that the model embeds two kinds of heterogeneity among workers, ex-ante and ex-post, it is interesting to provide some evidence on the relative importance of each in accounting for the positive cross-sectional correlation between gross inflow and outflow rates. To do so, I perform three experiments. First, I consider an economy in which all workers are ex-ante identical ($\theta = 0$), while keeping the other parameters the same as in Table 5. Second, I consider an economy in which all workers are ex-post identical ($p = 0$), while keeping the other parameters the same as in Table 5. Last, I also consider a version of the model in which there is no ex-ante heterogeneity ($\theta = 0$) and the model's parameters are estimated once again to match the same moments as in Section 4. Table 10 presents the correlation coefficients between gross outflow rates and net and gross inflow rates implied by these different versions of the model.[28]

[28] As before, the moments targeted in the estimation of the model are reported in bold. In all three cases, the cross-sectional correlation between gross and net inflow rates is positive and large.

Table 10
Comparisons Across Model Economies

	Cross-Sectional Correlations	
	Inflow-Outflow	Net Flow-Outflow
Data	0.56	**0.02**
Benchmark Model	0.65	**0.02**
No Ex-Ante Heterogeneity ($\theta = 0$ and constant parameters)	-0.15	-0.78
No Ex-Post Heterogeneity ($p = 0$ and constant parameters)	0.62	-0.14
No Ex-Ante Heterogeneity ($\theta = 0$ and re-estimated parameters)	0.39	**0.02**

The quantitative implications of the model with no ex-post heterogeneity are quite similar to those of the benchmark model. In contrast, the version of the model without ex-ante heterogeneity that keeps constant all other parameters produces correlations that are strongly counterfactual. These results are not entirely surprising given that ex-post heterogeneity accounts for only 16 percent of total migration. This estimate is in turn due to the fact that the fraction θ of workers who are ex-ante movers is set to target the fraction of young workers who never move in 3 model-periods. According to Kennan and Walker (2006), 82 percent of the individuals in their sample never move. This figure implies a relatively large value for θ and, conversely a relative low value for the parameter p. Lower values of the latter are associated with a smaller role for ex-post heterogeneity.

A version of the model in which there is no ex-ante heterogeneity ($\theta = 0$) *and* in which the remaining parameters are re-estimated along the lines described in Section 4, can account for the positive cross-sectional correlation between inflow and outflow rates (last row of Table 10). However, it tends to generate a value of this correlation that is smaller than in the data. The intuition why a positive amount of repeat movers ($\theta > 0$) helps to improve the model's fit in this dimension is simple. In the model with only ex-post heterogeneity, the positive inflow-outflow correlation relies on the fact that a fraction p of incoming workers decides to migrate again in the following period.[29] If $\theta > 0$, instead, a fraction $(\delta\theta/\overline{x} + p)$ of incoming workers decides to migrate again. This higher figure contributes to increase the cross-sectional correlation between inflow and outflow rates.

Last, while for tractability the modelling of ex-ante heterogeneity in the paper takes the extreme form of postulating a type that moves in every period, a plausible source of these ex-ante differences among workers is heterogeneity in moving costs. I leave this extension to future research. However, I conjecture that explicitly allowing for such form of heterogeneity would increase the quantitative importance of matching effects in accounting for the positive association between inflow and outflow rates. In such a version of the model, in fact, all workers moving to a new location would learn something about their idiosyncratic match.

[29] Notice that p is pinned down by the average migration rate in the economy, so it cannot be chosen to target the observed correlation between outflow and inflow rates.

Thereafter, poorly matched workers with relatively low costs of migration would migrate again. Without idiosyncratic matching effects the only source of migration in the model would be location-specific shocks, which fail to generate sizable migration flows.

8 Summary and Conclusions

This paper constructs a general equilibrium model of gross and net migration across locations to account for the fact that, in the data, U.S. states with relative high inflow rates of workers are also characterized by relatively high outflow rates. Standard models of employment reallocation across geographic areas cannot account for this pattern. Moreover, Census data on worker flows suggest that the latter occur within narrowly defined education, experience, industry cells. Therefore, explanations based on a changing composition of states' labor force cannot readily account for this fact, either. The model is estimated and can account for many of the cross-sectional and time-series facts about worker flows, and state-level labor earnings and rents.

The main mechanism emphasized in the theoretical model is unobserved heterogeneity among otherwise similar workers. The model allows for both ex-post differences among workers due to matching effects between a worker and a location, and ex-ante differences due to fixed workers' characteristics different from education and age. The quantitative results suggest that heterogeneity in fixed characteristics play an important role in accounting for the positive association between inflow and outflow rates across locations.

From a macro perspective, the main message of the paper is that cross-sectional heterogeneity in out-migration rates across U.S. states cannot be explained by the observable sources of heterogeneity among workers, such as age and education, traditionally emphasized in the migration literature. It also cannot be accounted for exclusively by aggregate shocks to locations, as emphasized in the macro-labor literatures. Instead, the model generates dispersion in outflow rates across states through the interaction of location-specific shocks and a mix of ex-ante and ex-post heterogeneity among workers. From this perspective, this paper has made progress toward bridging the gap between the mostly partial equilibrium literature on migration (e.g., Kennan and Walker, 2006) with the general equilibrium approach to employment reallocation across locations which typically does not allow for workers' idiosyncratic shocks (e.g. Lucas and Prescott, 1974).

References

Alvarez, Fernando and Marcelo Veracierto (2000): "Labor-Market Policies in an Equilibrium Search Model," in NBER Macroeconomics Annual, Volume 14. Cambridge and London: MIT Press, 265-304.

Alvarez, Fernando and Marcelo Veracierto (2006): "Fixed-Term Employment Contracts in an Equilibrium Search Model," NBER working paper series no. 12791.

Armenter, Roc and Francesc Ortega (2007): "Credible Redistributive Policies and Migration Across U.S. States," Working paper Federal Reserve Bank of New York.

Blanchard, Olivier and Lawrence Katz (1992): "Regional Evolutions," *Brooking Papers on Economic Activity* 1992, 1-75.

Card, David (2001): "Immigrant Inflows, Native Outflows, and the Local Market Impacts of Higher Immigration," *Journal of Labor Economics* 19, 22-64.

Caselli, Francesco and Wilbur Coleman (2001): "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation," *Journal of Political Economy* 109, 584-616.

Clark, Todd (1998): "Employment Fluctuations in U.S. Regions and Industries: The Roles of National, Region-Specific, and Industry-Specific Shocks," *Journal of Labor Economics*, 202-229.

Cobb, James (1993): The Selling of the South: the Southern Crusade for Industrial Development 1936-1990, University of Illinois Press, 2nd edition.

Duffie, Darrell and Kenneth Singleton (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica* 61, 929-952.

English, Beth (2006): A Common Thread: Labor, Politics, and Capital Mobility in the Textile Industry, Georgia University Press.

Flinn Christopher (1986): "Wages and Job Mobility of Young Workers," *Journal of Political Economy* 94, S88-S110.

Greenwood, Michael (1975): "Research on Internal Migration in the United States: A Survey," *Journal of Economic Literature* 13, 397-433.

Greenwood, Michael and Gary Hunt (1984): "Migration and Interregional Employment Redistribution in the United States," *American Economic Review* 74, 957-969.

Kennan, John and James Walker (2006): "The Effect of Expected Income on Individual Migration Decisions, " Working Paper, University of Wisconsin-Madison.

Kambourov, Gueorgui and Iourii Manovskii (2004): "Occupational Mobility and Wage Inequality," Working Paper, University of Pennsylvania.

Jovanovic, Boyan (1979): "Job Matching and the Theory of Turnover," *Journal of Political Economy* 87, 972-990.

Jovanovic, Boyan and Robert Moffitt (1990): "An Estimate of a Sectoral Model of Labor Mobility," *Journal of Political Economy* 98, 827-852.

Lkhagvasuren, Damba (2005): "Big Locational Differences in Unemployment Despite High Labor Mobility," Working Paper, University of Rochester.

Lee Bong-Soo and Beth Ingram (1991): "Simulation Estimation of Time-Series Models," *Journal of Econometrics* 47, 197-205.

Lucas, Robert and Edward Prescott (1974): "Equilibrium Search and Unemployment," *Journal of Economic Theory* 7, 188-209.

Miller, Ann (1967): "The Migration of Employed Persons to and from Metropolitan Areas of the United States," *Journal of the American Statistical Association* 62, 1418-1432.

Miller, Robert (1984): "Job Matching and Occupational Choice," *Journal of Political Economy* 92, 1086-1120.

Peet, Richard (1983): "Relations of Production and Relocation of United States Manufacturing Industry Since 1960," *Economic Geography* 59, 112–143.

Press W., Teukolsky, S., Vetterling, W., Flannery, B., 1996. Numerical Recipes in Fortran 77 (Cambridge University Press, Cambridge).

Rappaport J. (2004): "Why are Population Flows so Persistent?," *Journal of Urban Economics* 56, 554–580.

Roback, Jennifer (1982): "Wages, Rents and the Quality of Life," *Journal of Political Economy* 90, 1257-1278.

Steven Ruggles, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander (2004). Integrated Public Use Microdata Series: Version 3.0 [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor].

Sjaastad, Larry (1961): "Migration in the Upper MidWest," in *Four Papers on Methodology*, Upper Midwest Economic Study, Technical Paper no.1, 41-69.

Sjaastad, Larry (1962): "The Costs and Returns of Human Migration," *Journal of Political Economy* 70, 80-93.

Topel, Robert (1986): "Local Labor Markets," *Journal of Political Economy* 94 (Supplement), S111-43.

Treyz, George, Rickman, Dan, Hunt Gary and Michael Greenwood (1993): "The Dynamics of U.S. Internal Migration," *Review of Economics and Statistics* 75, 209-214.

Van Nieuwerburgh, Stjin and Pierre-Olivier Weill (2006): "Why Has House Price Dispersion Gone Up?" NBER working paper series no. 12538.

A Data Appendix

A.1 Sample Selection and Definitions

A.1.1 Sample Selection

Data are from the 5% samples of the 2000 Census, the 1990 Census (State Sample), the 1980 Census (State Sample), and from the 1% sample from the 1970 Census (Form 1 State Sample).[30] All the measures of gross and net flows and the stock of population that are reported in the paper are computed using a sample of individuals that, at the time of the relevant Census, satisfy the following restrictions:

- were between 27 and 60 years of age;
- were not living in group quarters;
- were in the labor force but not in the armed forces;
- if foreign-born, had immigrated to the U.S. at least 5 years before the Census year;
- were not living abroad 5 years before the Census year;
- were not living in the Census year or 5 years before the Census year, in either Alaska, Hawaii, or the District of Columbia.

A.1.2 Measures of Worker Flows

In order to construct measures of gross and net flows I adopt the following procedure. Individual i is observed living in state j in Census year τ. The same individual is also observed living in state k in year $\tau - 5$. Construct an indicator function $I_{i\tau}(j)$ for each individual i such that $I_{i\tau}(j) = 1$ if individual i was recorded as living in location j in Census year τ and zero otherwise. Also, define an indicator function $\overline{I}_{i\tau}(j)$ such that $\overline{I}_{i\tau}(j) = 1$ if individual i, interviewed in Census year τ, reported living in location j in year $\tau - 5$. Total outflow of population from location j between $\tau - 5$ and τ is then defined as

$$\text{outflows}_{j\tau} = \sum_{i} \sum_{k \neq j} \mu_{i\tau} \overline{I}_{i\tau}(j) \, I_{i\tau}(k) \, ,$$

where $\mu_{i\tau}$ is the person weight (`perwt`) assigned by the year τ Census to individual i. The total inflow of population into location j between $\tau - 5$ and τ is analogously defined as:

$$\text{inflows}_{j\tau} = \sum_{i} \sum_{k \neq j} \mu_{i\tau} \overline{I}_{i\tau}(k) \, I_{i\tau}(j) \, .$$

[30] Extending the analysis before 1970 presents some difficulty. The 1960 Census does not report a person's state of residence in 1955, but only if the person migrated across states or not. Thus, in 1960 it is only possible to compute gross inflows, but not gross outflows or net flows. In the 1950 Census, the migration question pertains to one year before, rather than 5 years before.

Let $\overline{y}_{j\tau}$ denote the total population that was interviewed in Census year τ and that was living in location j in year $\tau - 5$:

$$\overline{y}_{j\tau} = \sum_i \mu_{i\tau} \overline{I}_{i\tau}(j).$$

An outflow *rate* from location j between $\tau - 5$ and τ is then defined as follows:

$$\text{outflow rate}_{j\tau} = \frac{\text{out}_{j\tau}}{\overline{y}_{j\tau}}.$$

Analogously an inflow *rate* into location j between $\tau - 5$ and τ is defined as

$$\text{inflow rate}_{j\tau} = \frac{in_{j\tau}}{\overline{y}_{j\tau}}.$$

The net flow *rate* into location j between $\tau - 5$ and τ is defined as the difference between inflow and outflow rates. The absolute net flow rate for a location j is simply the absolute value of the net flow rate. Excess reallocation for a state j is defined as the sum of outflow and inflow rates minus the absolute net flow rate.

A.1.3 Demographic Groups

In order to control for demographic differences across states I construct 385 demographic groups based on the following variables (2000 Census):

- age (`age`); 7 age groups: 27-31, 32-36, 37-41, 42-46, 47-51, 52-56, 57-60;
- education (`educ99`); 5 education groups: high-school dropout, high-school diploma, some college, college degree, above college;
- industry of employment (`ind1990`); 11 industries: (1) agriculture, fishing, forestry, hunting and mining, (2) construction, (3) manufacturing non-durables, (4) manufacturing durables, (5) transportation, communication and other public utilities, (6) wholesale and retail trade, (7) finance, insurance, and real estate, (8) business and repair services, (9) personal services, entertainment and recreation services, (10) professional and related services, (11) public administration.

Denote each cell by g and the collection of cells by G. There are 385 cells. For each cell g it is possible to construct the equivalents of total outflows, inflows, and population defined above in the following way:

$$\begin{aligned}
\text{outflows}_{jg\tau} &= \sum_{i \in g} \sum_{k \neq j} \mu_{i\tau} \overline{I}_{i\tau}(j) I_{i\tau}(k), \\
\text{inflows}_{jg\tau} &= \sum_{i \in g} \sum_{k \neq j} \mu_{i\tau} \overline{I}_{i\tau}(k) I_{i\tau}(j), \\
\overline{y}_{jg\tau} &= \sum_{i \in g} \mu_{i\tau} \overline{I}_{i\tau}(j).
\end{aligned}$$

Group-specific outflow and inflow rates are then defined as

$$\begin{aligned} \text{outflow rate}_{jg\tau} &= \frac{\text{outflows}_{jg\tau}}{\overline{y}_{jg\tau}}, \\ \text{inflow rate}_{jg\tau} &= \frac{\text{inflows}_{jg\tau}}{\overline{y}_{jg\tau}}. \end{aligned}$$

A.1.4 Weekly Earnings

Workers' weekly earnings were computed using data from the Census of Population and Housing. The same sample selection criteria listed above in Section A.1.1 were also applied in this case. Weekly earnings were obtained by summing, for each worker, annual wage income (`incwage`) and business and farm income (`incbus+incfarm`), and dividing the sum by the number of weeks worked (`wkswork1`). Each source of income refers to the year preceding the Census year. I have dropped from the sample a very small number of observations for which an individual reported zero annual earnings but a positive number of weeks worked. In a few instances, reported earnings by self-employed individuals were negative, and these observations have been dropped as well. Given that earnings refer to the year prior to the Census and the worker's labor force participation status refers to the time of the survey, a small fraction of individuals (about 2.5 percent of the sample) reported zero annual earnings and zero weeks worked in the year prior to the Census. I have also dropped these individuals from the sample.

For each Census year, the logarithm of weekly earnings was regressed on the following variables: 48 dummies for workers' state of residence in the Census year (`statefip`), a measure of workers' experience (computed subtracting years of education from the workers' age) and experience squared, 17 education dummies (`educ99`), a workers' sex (`sex`), 3 race dummies ("white", "black" and "others", constructed from `raced`), 11 sectoral dummies (constructed from `ind1990`), and 6 occupational dummies (constructed from `occ1990`). The R^2 of these regressions was typically 30 percent.

The measure of average weekly earnings for each state is represented by the estimates of state fixed effects in this regression.

A.1.5 Rents

The Census of Population and Housing provides data on the gross monthly rent (`rentgrs`) paid by a renter. This variable reports the gross monthly rental cost of the housing unit, including contract rent plus additional costs for utilities (water, electricity, gas) and fuels (oil, coal, kerosene, wood, etc.). This information is used to derive a measure of land rents in each U.S. state. Observations on rents were obtained for those workers renting a housing units who satisfy the sample selection criteria listed above in Section A.1.1. In each Census year, about 30 percent of the sample obtained from applying those selection criteria rents (as opposed to owns) its housing unit. For example, in the 2000 Census more than one million observations are available for renters.

In order to remove the influence of observable characteristics of a housing unit from the monthly rent, I ran an hedonic regression of the logarithm of the rent on state fixed effects and

a list of observable characteristics of the housing units. These include: a dummy for whether the housing unit is located in a metropolitan area (`metro`), a dummy for whether the unit is used commercially (`commuse`), a dummy about the acreage of the property (`acreprop`), a dummy about the acreage of the house (`acrehous`), a dummy on whether meals are included in the rent (`rentmeal`), a dummy for whether the housing unit is in a condominium (`condo`), a dummy on whether the housing unit contains a kitchen (`kitchen`), a dummy on the number of rooms (`rooms`), a dummy about the availability of plumbing facilities (`plumbing`), a dummy about the age of the unit (`builtyr`), a dummy about the number of bedrooms (`bedrooms`), a dummy about the number of units in the structure (`unitsstr`).

The measure of average rents at the state level is represented by the estimated state fixed effects from this regression.

A.2 Immigration

The sample of workers selected according to the criteria spelled out in Section A.1.1 includes foreign-born workers provided they migrated to the U.S. at least 5 years before the Census year. In what follows I refer to these workers as "recent" immigrants. While this selection guarantees that aggregate net flows of workers across U.S. states are zero in each Census year, it also assumes away recent immigration flows. The latter might potentially play an important role in affecting internal migration flows from and into certain states. In order to quantify these effects and to place the magnitude of internal migration flows in perspective, relative to their international counterpart, I have used the 2000 Census data to compute for each U.S. state the ratio between the number of recent immigrants that were located in that state in the year 2000 and the state's 1995 population. The data indicate that the average rate of recent immigration across U.S. states is about 1.5 percent, while Table 1 indicates that the average inflow rate of workers from the rest of the U.S. in the 2000 Census is about 8.6 percent. Thus, for the average state internal migration is much larger than recent immigration. Of course, it is well known that recent immigrants tend to cluster in a few locations. There are three states for which the recent immigration rate is at least 30 percent as large as the gross inflow rate due to internal migration: California (47 percent), New York (56 percent), and New Jersey (34 percent). It turns out, however, that for the cross-section of U.S. states there is no significant association between rates of recent immigration and both gross and net flow rates due to internal migration. So, there is no evidence supporting the view that states with low gross inflow rates experienced larger than average immigration from abroad. This result is consistent with the findings of Card (2001) for U.S. metropolitan areas.

B The Benchmark Model

Under the assumption that the shock υ_1 is sufficiently large, workers whose match is bad will always choose to migrate. Their value function is then given by:

$$V(s,\zeta,\upsilon_1) = \max_{c,l^c} \left\{ c + \phi(l^c) - \upsilon_1 + \beta\delta(e-k) + \beta(1-\delta)e_r \right\}. \tag{17}$$

The first order condition for land is:

$$r(s,\zeta) = \phi'(l^c).$$

This equation can be solved for the optimal l^c, denoted by $l^c(s,\zeta)$, which can then be plugged back into the value function:

$$V(s,\zeta,\upsilon_1) = u(s,\zeta,\upsilon_1) + \beta\delta(e-k) + \beta(1-\delta)e_r,$$

where, for convenience, I have defined the following indirect utility function for agents which draw a shock υ:

$$u(s,\zeta,\upsilon) = w(s,\zeta) - r(s,\zeta)\,l^c(s,\zeta) + \phi(l^c(s,\zeta)) - \upsilon.$$

It is possible to simplify the expression for $u(s,\zeta,\upsilon)$ using the functional form (14) for ϕ (with $\alpha = 1-\tau$) and using the fact that due to the constant returns to scale assumption unit cost of production must equal one (the price of output). The latter equation, which can be easily obtained by manipulating the firms' first order condition, can then be used to solve for $r(s,\zeta)$ as a function of $w(s,\zeta)$ and the shock z:

$$r(s,\zeta) = \left[\frac{z(1-\tau)^{1-\tau}(\tau)^{\tau}}{w(s,\zeta)^{\tau}}\right]^{\frac{1}{1-\tau}}. \tag{18}$$

The indirect utility function $u(s,\zeta,\upsilon)$ in this case becomes:

$$u(s,\zeta,\upsilon) = w(s,\zeta)\left(1 + z^{-\frac{1}{\tau}}\left(\frac{A}{1-\tau}\right)^{\frac{1}{\tau}}\right) - \upsilon. \tag{19}$$

Land can be completely solved for by using the equilibrium conditions (11) and (12) together with the worker's first order condition for land (10). Replace the latter into equation (12) to get:

$$y\left(\frac{A}{r(s,\zeta)}\right)^{\frac{1}{1-\alpha}} + l^c(s,\zeta) = 1.$$

Now, solve for $l^c(s,\zeta)$ and replace it in the inverse demand function for labor:

$$w(s,\zeta) = \tau z y^{\tau-1}\left(1 - y\left(\frac{A}{r(s,\zeta)}\right)^{\frac{1}{1-\alpha}}\right)^{1-\tau}.$$

Use (18) to replace $r(s,\zeta)$ on the right-hand side of this equation. This yields an equation with $w(s,\zeta)$ on both the right and left-hand sides. The advantage of the assumption $\alpha = 1-\tau$ is that this equation can be solved in closed-form to yield the wage as a function of the state variables of the model:

$$w(s,\zeta) = \frac{\tau z y^{\tau-1}}{\left[1 + z^{-\frac{1}{\tau}}\left(\frac{A}{1-\tau}\right)^{\frac{1}{\tau}}\right]^{1-\tau}}. \tag{20}$$

I have then shown that it is possible to solve out for $r(s,\zeta)$ and land from the original specification and rewrite the model as one in which the utility functions take the form (19) and the unit wage is given by (20).

With these preliminary steps, the model's equilibrium can then be characterized as follows. Given the calibration of the model, workers who draw the idiosyncratic shock $\upsilon = \upsilon_1$ always choose to move, so that $q(s,\zeta,\upsilon_1) = 1$. This fact allows me to merge the state variables n and m into a new state variable

$$\overline{n} = n + m,$$

which includes both workers who belong to a dynasty that always relocates exogenously and workers who draw a low idiosyncratic shock $\upsilon = \upsilon_1$. The state vector s is now comprised of the following variables:

$$s = (y, \overline{n}, \zeta).$$

It is then possible to characterize the equilibrium of the model in relation to the behavior of workers for whom $\upsilon = \upsilon_2$. Following Lucas and Prescott (1974), we need to distinguish among three different regions:

- **Region 1**. Some (or all) of these workers choose to leave and some (or none) choose to remain. In this case, the value of staying must be less or equal to the value of leaving, with strict inequality prevailing if everybody chooses to leave:

$$\int V(s',\zeta',\upsilon_2)\, Q(\zeta, d\zeta') \leq e - k. \tag{21}$$

  Notice that in this case, endogenous inflows are zero, $x_d(s,\zeta) = 0$, because workers with a high location shock choose to leave and the expected location shock of incoming workers would be strictly less than υ_2. Thus, the components of s' are:

$$\begin{aligned} Y(s,\zeta) &= \delta(y-n)(1-q(s,\zeta,\upsilon_2)) + y(1-\delta+\overline{x}(1-\eta)), \\ \overline{N}(s,\zeta) &= (p(1-\delta\theta/\overline{x}) + \delta\theta/\overline{x})\, y\overline{x}(1-\eta) + (p(1-\theta)+\theta)\, y(1-\delta) \end{aligned}$$

Notice that $q(s,\zeta,\upsilon_2) \leq 1$ is implicitly defined by (21) in case of equality.

- **Region 2**. None of the υ_2 workers chooses to leave and no new worker chooses to locate there. In this case:

$$\int V(s',\zeta',\upsilon_2)\, Q(\zeta, d\zeta') > e - k, \tag{22}$$

$$p\int V(s',\zeta',\upsilon_1)\, Q(\zeta, d\zeta') + (1-p)\int V(s',\zeta',\upsilon_2)\, Q(\zeta, d\zeta') < e_d. \tag{23}$$

  The first inequality expresses the fact that it is better for a υ_2 type of worker to remain in the location, while the second inequality expresses the fact that no migrating worker will choose to migrate to this location. The endogenous components of s' are:

$$\begin{aligned} Y(s,\zeta) &= \delta(y-n) + y(1-\delta+\overline{x}(1-\eta)), \\ \overline{N}(s,\zeta) &= (p(1-\delta\theta/\overline{x}) + \delta\theta/\overline{x})\, y\overline{x}(1-\eta) + (p(1-\theta)+\theta)\, y(1-\delta). \end{aligned}$$

Notice that, using equation (17), the two inequalities (22) and (23) can be rewritten as:[31]

$$e-k < \int V\left(s',\zeta',\upsilon_2\right)Q(\zeta,d\zeta') < \frac{e_d - p\left(\int u(s',\zeta',\upsilon_1)Q(\zeta,d\zeta') + \beta\delta\left(e-k\right) + \beta\left(1-\delta\right)e_r\right)}{1-p}.$$

- **Region 3**. None of the υ_2 workers chooses to leave and some new workers choose to locate there. In this case, equation (22) still holds, while equation (23) holds as an equality:

$$e-k < \int V\left(s',\zeta',\upsilon_2\right)Q(\zeta,d\zeta') = \frac{e_d - p\left(\int u(s',\zeta',\upsilon_1)Q(\zeta,d\zeta') + \beta\delta\left(e-k\right) + \beta\left(1-\delta\right)e_r\right)}{1-p}. \tag{24}$$

The endogenous components of s' are:

$$\begin{aligned} Y\left(s,\zeta\right) &= x_d(\zeta,s) + \delta\left(y-n\right) + y\left(1-\delta+\overline{x}\left(1-\eta\right)\right), \\ \overline{N}\left(s,\zeta\right) &= \left(p\left(1-\delta\theta/\overline{x}\right)+\delta\theta/\overline{x}\right)\left(x_d(\zeta,s)+y\overline{x}\left(1-\eta\right)\right) + \left(p\left(1-\theta\right)+\theta\right)y\left(1-\delta\right). \end{aligned}$$

where $x_d(s,\zeta)$ is implicitly defined by the equality condition in (24).

C Extension: Amenities in the Utility Function

It is easy to modify the model to include location-specific amenities and to show that an appropriate choice of functional forms yields exactly the same implications for net and gross flows of labor as the one derived in the previous section. Suppose that there are no location-specific productivity shocks z, but rather that each location is characterized by a time-varying amenity a. The utility function takes the form:

$$u = h\left(a\right)\left(c+\phi\left(l^c\right)\right) - \upsilon,$$

where $h(.)$ is an increasing function of a.

Solving for the optimal choice of land and consumption yields the following indirect utility function:

$$u\left(s,\widehat{\zeta},\upsilon\right) = h\left(a\right)\left(w\left(s,\widehat{\zeta}\right) - r\left(s,\widehat{\zeta}\right)l^c\left(s,\widehat{\zeta}\right) + \phi\left(l^c\left(s,\widehat{\zeta}\right)\right)\right) - \upsilon,$$

where $\widehat{\zeta}$ denotes the vector (a,ε).

Replacing the expressions for the wage and rent yields:

$$u\left(s,\widehat{\zeta},\upsilon\right) = h\left(a\right)w\left(s,\widehat{\zeta}\right)\left(1+\left(\frac{A}{1-\tau}\right)^{\frac{1}{\tau}}\right) - \upsilon.$$

[31] It is straightforward to check that the condition $\beta\delta < 1$ guarantees that the left-most term is smaller than the right-most term in this equation.

Since wages are given by:

$$w\left(s,\widehat{\zeta}\right)=\frac{\tau y^{\tau-1}}{\left[1+\left(\frac{A}{1-\tau}\right)^{\frac{1}{\tau}}\right]^{1-\tau}},$$

I obtain that

$$u\left(s,\widehat{\zeta},\upsilon\right)=h\left(a\right)\tau y^{\tau-1}\left(1+\left(\frac{A}{1-\tau}\right)^{\frac{1}{\tau}}\right)^{\tau}-\upsilon.$$

This model has the same reduced-form utility function as the benchmark model as long as $h(a)$ takes the functional form:

$$h(a)=\frac{\left(a^{\frac{1}{\tau}}+\left(\frac{A}{1-\tau}\right)^{\frac{1}{\tau}}\right)^{\tau}}{\left(1+\left(\frac{A}{1-\tau}\right)^{\frac{1}{\tau}}\right)^{\tau}},$$

and a follows the same stochastic process as z:

$$\begin{aligned} a' &= a\varepsilon', \\ \varepsilon' &= a^{\psi-1}\varepsilon^{\rho}u'. \end{aligned}$$

Notice that the model with amenities has the same implications for flows of workers across states, but has different implications for prices of labor and land. Specifically, rents would tend to be higher and wages lower in locations with better amenities (as in Roback, 1982).

D Details On Numerical Implementation

This section describes the steps that I followed in solving and estimating the model. The algorithm is comprised of three loops: one for finding the value function conditional on $\overline{x}$, e_d, e_r, and the parameter vector ϕ, one for finding the equilibrium values $(\overline{x}, e_d, e_r)$, of the model for given ϕ, and one for finding ϕ in order to match the empirical moments of interest. Every change in ϕ entails new equilibrium values for $(\overline{x}, e_d, e_r)$, while a new guess for $(\overline{x}, e_d, e_r)$ requires the computation of the associated value function.

Step 1 (Guess). Start from an initial guess for the parameter vector ϕ and for the values of directed and undirected search e_d and e_r. The guess for $(\overline{x}, e_d, e_r)$ is updated in Step 3 below, while the guess for ϕ is updated in Step 4.

Step 2 (Dynamic Programming). Solve the dynamic programming problem for workers such that $\upsilon = \upsilon_2$. This is the most time-consuming step of the procedure because there are four continuous state variables in the problem (recall that s includes y and $\overline{n}$ while z includes ζ and ε) and because the procedure involves numerical integration of the value function with respect to the density of the innovation u. The solution of the dynamic programming problem yields gross inflows $x(s,\zeta)$ and the probability of outflow $q(s,\zeta,\upsilon_2)$ for an agent with match υ_2 as functions of the state vector (s,ζ). These two functions allow one to recover all the other variables of interest conditional on (s,ζ).

Step 3 (Equilibrium). Solve for the equilibrium value of $(\overline{x}, e_d, e_r)$ by defining the function $f(\overline{x}, e_d, e_r)$ with the following components:

$$f_1(\overline{x}, e_d, e_r) = \int x(s,\zeta)\mu(ds, d\zeta) - \eta\overline{x}, \tag{25}$$

$$f_2(\overline{x}, e_d, e_r) = \int y\mu(ds, d\zeta) - 1, \tag{26}$$

$$f_3(\overline{x}, e_d, e_r) = p\int V(s,\zeta,\upsilon_1)\, y\mu(ds, d\zeta) + (1-p)\int V(s,\zeta,\upsilon_2)\, y\mu(ds, d\zeta) - e_r. \tag{27}$$

The integral in equations (25)-(27) are computed by simulating the economy for a very large number of periods (5 million), obtaining $\{x_t, y_t, V_{1t}, V_{2t}\}_{t=1}^{T}$ and approximating the integrals in (25)-(27) with the corresponding sample averages. For example, the integral in (26) is approximated by:

$$\frac{1}{T}\sum_{t=1}^{T} y_t.$$

The vector $(\overline{x}^*, e_d^*, e_r^*)$ such that $f(\overline{x}^*, e_d^*, e_r^*) = 0$ is computed using Broyden's algorithm described in the Step 4. Notice that for each candidate value of $(\overline{x}, e_d, e_r)$ it is necessary to go back to Step 2 and solve the dynamic programming problem again.

Step 4 (Estimation). Given $(\overline{x}^*, e_d^*, e_r^*)$, it is feasible to compute the equilibrium value of all the variables of interest. The vector ϕ is estimated by constructing the model counterpart of the five moments listed in the text (Section 4) and choosing ϕ so that the model-generated moments are exactly equal to their empirical counterparts. Since there are five parameters and five moments, this is an exactly identified model. The problem then becomes one of solving five non-linear equations in five unknowns. The model-generated moments are constructed by using 5 million simulated data drawn from the model. Each simulated moment is then compared with its empirical counterpart. In order to find a solution for this non-linear system of five equations in five unknowns I have used Broyden's algorithm. The latter operates in the following way (for a more detailed description, see Press et al., 1996, chapter 9). First, it numerically approximates the Jacobian matrix associated with the non-linear system. It then uses this approximate Jacobian to find an updated vector ϕ by implementing the Newton step, which guarantees quadratic convergence if the initial guess is close to the solution. If the Newton step is not "successful", the algorithm tries a smaller step by back-tracking along the Newton dimension. When an acceptable step is determined, ϕ is updated and the algorithm can proceed in the way described above, once an updated Jacobian has been obtained. Since the numerical computation of the Jacobian can be costly (and in this model it is), the Jacobian at the new vector ϕ is iteratively approximated using Broyden's formula. The non-linear solver stops when the maximum percentage difference between the simulated moments and the empirical moments is, in absolute value, smaller than 10^{-4}.