# Understanding Gross Worker Flows Across U.S. States

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

This Version: November 2007. First Version: February 2006.
Comments welcome.

**Abstract**

Migration of workers across states is the main factor behind the large and persistent dispersion in the growth rates of employment across U.S. states (Blanchard and Katz (1992)). The analysis of gross flows of workers helps us understand the mechanisms by which variations in net workers flows occur. This paper has two goals. The first one is to document in a systematic way the main stylized facts about net and gross flows of workers across states, state-level earnings, and rents using micro data from the U.S. Census of Population and Housing. The second goal is to provide an explanation and an interpretation of these facts using a dynamic general equilibrium model of workers' migration. The calibrated model is successful in accounting for most of the documented stylized facts. The model suggests that the response of a state's average wage to a shock that reduces employment in the state depends on whether the adjustment process works through an increase in gross outflows of workers or a drop in gross inflows. This implication is consistent with time-series variation in state-level earnings across Census years.

*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, UT-Austin, SUNY Albany, Cleveland Fed, 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, and Ted Temzelides for useful discussions and suggestions. Richard Lowery provided able research assistance.

# 1 Introduction

Migration of workers across states is the main factor behind the large and persistent dispersion in the growth rates of employment across U.S. states (Blanchard and Katz (1992)). Yet, the process by which population is reallocated among geographic areas within a country is not well understood yet. This paper argues that in order to improve this understanding it is important to consider *both* gross and net flows of workers across states. The analysis of both kinds of flows allows one to determine the extent to which net gains of employment by some states are due to higher gross inflows or, alternatively, to lower gross outflows of workers.

This paper has two goals. The first one is to construct empirical measures of net and gross flows of workers across states and to characterize the main cross-sectional and time-series stylized facts in this area using data from the U.S. Census of Population and Housing. The second goal is to show that these stylized facts can be explained using a dynamic general equilibrium model of workers' and firms' flows across states.

I start by documenting the key facts on workers' migration using the decennial Census of the U.S. for the period 1970-2000. The latter allows one to determine a respondent's state of residence in the Census year as well as five years before the Census year. This information is used to construct state-level aggregate gross and net rates of workers flows. The key stylized facts are as follows. First, gross flows of workers are large relative to net flows. For example, between 1995 and 2000 the average state gained or lost about 2.1 percent of its 1995 population. In the same period, the average state experienced a combined inflow and outflow of population of about 17 percent of its 1995 population. Second, most interstate flows of workers occur within narrowly defined demographic groups. Third, in the cross-sectional dimension: (1) Gross inflow rates are more dispersed than net inflow rates, which are more dispersed than gross outflow rates. (2) Gross inflow and outflow rates are positively correlated. (3) Gross and net inflow rates are highly positively correlated, while net flow rates and gross outflow rates display a U-shaped relationship. Fourth, in the time-series dimension, there is a large degree of persistence in both gross and net flow rates across Census years for a given state. Fifth and last, average state weekly earnings and rents are approximately uncorrelated with net flow rates in the cross-section.

I interpret these stylized facts using an equilibrium model of gross and net migration flows. The model economy is composed by a set of local labor and land markets ("islands"), that are hit by shocks to total factor productivity (Lucas and Prescott (1974)). Ex-ante identical workers are assumed to be mobile across islands while firms use constant-returns-to scale technologies in labor and land to produce an homogeneous good (Roback (1982)). At a point in time, a location typically experiences both gross inflows and gross outflows of workers because of an idiosyncratic shock to the match between a worker and that location. In general equilibrium, the value of migration is pinned down by the requirement that aggregate net flows are zero. The model's parameters are estimated using a method of simulated moments. The estimated model can account for most of the stylized facts mentioned above.

In the model, shocks to the growth rate of total factor productivity induce persistent differences in net flow rates across locations. The extent to which variations in gross inflows or gross outflows account for variations in net flows depends on whether net flows are positive or negative. Locations that are net-receivers of population experience relatively large gross inflows of workers who are expected to arbitrage away the increase in local nominal wages. Outflow from these locations is then driven by selection effects: newly arrived workers exhibit larger turnover than incumbent ones, who are characterized by relatively better idiosyncratic matches with these locations. Thus, locations characterized by large and positive net flow rates also tend to exhibit large gross outflow rates. Hence the positive correlation between gross inflow and gross outflow rates. In locations that are net-losers of population adjustment occurs through lower gross inflows, rather than larger gross outflows, as long as the former flows are positive. When gross inflows hit their lower bound, the negative net flow is accommodated by higher gross outflows. The fact that gross outflows operate to accommodate negative net flows only in this extreme circumstance explains why gross outflows are less volatile than both gross and net inflows. Thus, the model suggests that reallocation of population within the U.S. occurs mainly through variations in gross inflows (large in fast-growing states and small in slow-growing states), rather than in gross outflows. In other words, states that tend to lose population to other states do so by attracting fewer new workers as opposed to losing more local ones.

This paper is related to several literatures. The closest literature is the one initiated by Lucas and Prescott (1974) with their "island" model of the labor market.[1] Lucas and Prescott develop a model of workers' net flows across locations driven by shocks to local labor demand. In a sense the present paper can be thought of as a version of Lucas and Prescott (1974) in which also workers are hit by idiosyncratic location-specific shocks, giving rise to gross flows of workers. The importance of gross flows of workers across sectors is highlighted by Jovanovic and Moffitt (1990) who consider a simplified version of the Lucas-Prescott model allowing for idiosyncratic sector-specific shocks to workers' productivity. An important insight of their model is that the introduction of idiosyncratic shocks has implications for the dynamics of sectoral wages. For example, in the Jovanovic and Moffitt paper net flows are such that unit wages are always equalized across sectors. In the original contribution by Lucas and Prescott, instead, the fact that workers are homogeneous within an island implies that equilibrium wage differentials across islands are necessary to give rise to net flows. In addition to focusing on geographic, as opposed to sectoral, mobility, my paper differs from Jovanovic and Moffitt (1990) in several important dimensions. First, Jovanovic and Moffitt do not estimate the parameters of their structural model, but rather tested some of its empirical implications. Second, 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 it is ill suited to the empirical application of the model. Third, Jovanovic and Moffitt focus on an equilibrium in which gross inflows into each sector are always strictly positive, so that unit wages are equalized across sectors. While this assumption greatly simplifies the analysis, its validity is an empirical issue. It turns out that, in my model, equilibria in which unit wages are equalized occur only under parameterizations of the model that are not consistent with the statistical properties of net flows 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 gross

[1] Topel (1986) considers a setting similar to Lucas and Prescott (1974). Variations of the Lucas-Prescott model have been used to analyze a host of labor market issues. Kambourov and Manovskii (2004) study trends in occupational mobility and wage inequality. Alvarez and Veracierto (2000 and 2006) analyze labor market policies.

inflows are occasionally zero.

Second, the paper is related to Roback (1982)'s classic model of the effect of local amenities on local land prices and wages. In fact, the model represents a generalization of Roback's model to a dynamic setting in which local amenities enter in either the production function or in the utility function and evolve stochastically over time, giving rise to the dynamics of local employment. 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 on the dynamics of workers flows rather than on housing prices.

Third, the paper also builds on the seminal contribution of Blanchard and Katz (1992), who develop a reduced form model of workers' net flows across U.S. states, and provide some interesting VAR evidence on the nature of states' adjustment process to local labor demand shocks. Relative to Blanchard and Katz, this paper also focuses on gross flows of workers and not only on net flows.[2]

Fourth, 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) and Treyz et al. (1993). The contribution of this paper relatively to this mostly empirical literature is to develop a tractable structural model of gross and net workers' flows.

Finally, this paper is related to the partial equilibrium literature on the determinants of workers' migration decisions. Kennan and Walker (2005) carefully estimate such model using NLSY data and use the panel structure of the data to identify wage differences due to location effects.[3]

The rest of the paper is organized as follows. Section 2 describes the data and the stylized facts. Section 3 presents the model. Section 4 describes the estimation of the model, discusses the fit, and performs sensitivity analysis. Section 5 discusses the implications of the model for the correlation between local wages and net flow rates. Section 6 concludes. The appendix offers 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.

[2] See also Barro and Sala-i-Martin (1992) and (1991), who, using state-level data dating back to 1840, study convergence in income per capita across U.S. states. More recent contributions by macroeconomists to the literature on internal migration of workers include Hassler et al. (2005) and Lkhagvasuren (2005). The former argue that differences in the generosity of unemployment insurance between the U.S. and Europe can explain higher internal mobility rates in the U.S. The latter paper tries to explain the existence of persistent differentials in unemployment rates across U.S. states by means of a general equilibrium matching model with location-specific idiosyncratic productivity shocks.

[3] The approach followed by Kennan and Walker (2005), and myself, to model migration is based on the idea that matching effects are important in generating workers' gross flows. This feature connects this paper to Miller (1984) and Flinn (1986)'s matching models based on Jovanovic (1979). Differently from these papers, both Kennan and Walker (2005) and I abstract from learning effects: a worker is assumed to learn about his match with a location upon arriving there.

# 2 Data and Stylized Facts

In this section I briefly describe the data used in the paper and then organize the main features of workers' 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)).[4]

Since 1940, the Census questionnaire has included 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.[5] 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 individuals as well as foreign-born ones who immigrated to the U.S. at least five years prior to the Census year. This restriction is necessary for aggregate net flows of workers to equal zero. Appendix A contains more detailed information on issues of sample selection as well as on the construction of the variables described below. From now on, for simplicity, I will refer to a state's "population" as the collection of individuals satisfying the sample selection criteria described in Appendix A.[6]

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 already short time-series dimension of the data. Second, about 20 percent of the U.S. population does not currently live in a

[4] This is available online at `http://usa.ipums.org/usa/`. The Census data have the clear advantage of being a large and comprehensive data set. Information on geographic mobility of individuals is available from other sources. For example, the March Current Population Survey (March CPS) contains such information, but only includes approximately 60,000 households. Given that on average only 3 percent of the population leaves its state of residence in a given year, this amounts to observing less than 2,000 households migrating across state lines, or, on average, 40 households per state. In contrast, the decennial Census typically contains information on million of households.

[5] The levels of inflow and outflow of population for a given state were standardized by the number of individuals 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.

[6] It is important to notice that the micro unit of analysis in this paper is an individual and not a household. For an analysis of the family migration decision see Mincer (1978).

metropolitan area. This figure has increased by about 10 percentage points since 1970, and it displays a non-trivial geographic variation. Therefore, also in this case there would be some ambiguity associated with the definition of a local labor market. 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. Fourth, and last, for the purpose of policy analysis, many labor market policies are set at the state level (see Armenter and Ortega (2007)).

## 2.2 Stylized Facts

In what follows I summarize the main features of the data on workers' flows in a few stylized facts using simple descriptive statistics. The next subsection discusses a few issues related to the interpretation of these statistics.

**Stylized fact #1.** *Gross workers' flows are large relative to net flows (Table 1).*

For example, between 1995 and 2000 the average state gained or lost about 2.1 percent of its 1995 population. In the same period, the average state experienced a combined inflow and outflow of population of about 17 percent of its 1995 population. The statistics in this and in the following tables are computed weighting each state by its relative population in order to avoid assigning disproportional importance to small outlier states.

Table 1
Basic Statistics on Workers Flows (Censuses 1970-2000)

| | 2000 | | 1990 | | 1980 | | 1970 | |
|---|---|---|---|---|---|---|---|---|
| | Mean | St. Dev. | Mean | St. Dev. | Mean | St. Dev. | Mean | St. Dev. |
| Outflow Rate (*out*) | 8.61 | 1.72 | 9.04 | 2.52 | 8.97 | 2.07 | 6.89 | 2.28 |
| Inflow Rate (*in*) | 8.61 | 3.75 | 9.04 | 4.19 | 8.97 | 4.80 | 6.89 | 3.35 |
| Net Flow Rate (*net*) | 0.00 | 2.86 | 0.00 | 3.99 | 0.00 | 3.85 | 0.00 | 2.54 |

**Stylized fact #2.** *In the cross-section, gross inflow rates are more dispersed than net inflow rates, and the latter are more dispersed than gross outflow rates (Table 1).*

From the cross-sectional standard deviations of net and gross flows it is feasible to deduce their cross-correlation patterns. These are reported in Table 2.

Table 2
Cross-Sectional Correlations (Censuses 1970-2000)

| | Outflow Rate | Inflow Rate | Net Flow Rate |
|---|---|---|---|
| Outflow Rate | | | |
| 1970 | 1 | 0.65*** | −0.04 |
| 1980 | 1 | 0.63*** | 0.24 |
| 1990 | 1 | 0.38** | −0.23 |
| 2000 | 1 | 0.69*** | 0.30** |
| Inflow Rate | | | |
| 1970 | | 1 | 0.73*** |
| 1980 | | 1 | 0.91*** |
| 1990 | | 1 | 0.81*** |
| 2000 | | 1 | 0.90*** |

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

These main features of these correlation patterns are summarized in stylized fact #3.[7] Figures 1-3 report scatter plots of outflow, inflow and net flow rates across U.S. states using the 2000 Census data.

**Stylized Fact #3.** *In the cross-section, gross and net inflow rates are strongly positively correlated (Figure 2). Gross inflow and outflow rates are positively correlated (Figure 1). Gross outflow rates and net flow rates tend to be uncorrelated (Figure 3). Moreover, the relationship between net flow rates and outflow rates is U-shaped.*

On this latter point, the small (in absolute value) correlations between outflow and net flow rates hide the effect of two opposing forces. As shown in Figure 3, for example, North Dakota and Nevada have both very large outflow rates (around 13-14%), even if they are at the opposite extremes of the distribution of net flows. This observation generalizes to the whole sample of states and different Census years. While in a regression of outflow rates on net flow rates the coefficient on net flows is typically insignificant, a similar regression with squared net flow rates on the right-hand side yields strongly significant positive coefficients.

**Stylized fact #4.** *Both gross and net flow rates are very persistent over time (Table 3).*

Table 3 reports, for each type of flow, its autocorrelation coefficient across Census years, computed by pooling all state-year data points together.[8]

[7] Notice that stylized fact #3 contains additional information relative to stylized fact #2 because it provides a ranking of *correlation coefficients* among the different flow rates. The latter cannot be inferred from the ranking of their variances.

[8] In order to increase the sample size in the time-series dimension, I have included the 1950 Census in these computations. The 1950 Census asked respondents to report their state of residence in 1949, as opposed to 1945. The 1 year migration data were converted into 5 year migration data by multiplying the 1 year flows by 3.79. This number is obtained by comparing the 1 year interstate migration rate from the 2000 American Community Survey with the 5 year migration rate from the 2000 Census. Notice that, as mentioned in footnote (26), the 1960 Census does not allow one to compute gross outflows and net flows, but only gross inflows. For this reason it cannot be used to construct table 3.

Table 3
Autocorrelations of Workers Flows (Censuses 1950, 1970-2000)

| | Autocorrelation Coefficient | | | |
|---|---|---|---|---|
| | $t, t-1$ | $t, t-2$ | $t, t-3$ | $t, t-4$ |
| Outflow Rate | 0.76*** | 0.79*** | 0.67*** | 0.60*** |
| Inflow Rate | 0.88*** | 0.83*** | 0.84*** | 0.68*** |
| Net Flow Rate | 0.58*** | 0.61*** | 0.53*** | 0.33** |

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

A more complete picture of the process of internal migration in the U.S. requires an analysis of measures of state-level labor earnings and land rents and how these are cross-sectionally correlated and autocorrelated over time. State-level labor earnings and land rents are measured as state fixed-effects in hedonic regressions of earnings and rents on observable characteristics of workers and renter-occupied housing units (see Appendix A for details.) The following stylized fact summarizes the relevant information.

**Stylized fact #5.** *In the cross-section, average weekly earnings and rents are positively correlated. The net flow rate tends to be uncorrelated with a state's average weekly earnings and average rents. Average rents are more dispersed than average weekly earnings. In the time-series dimension, the ranking of states with regard to both average weekly earnings and rents is persistent across Census years (Table 4).*

Table 4
Earnings, Rents and Net Flows (Censuses 1970-2000)

| | Correlation with | | Standard | Spearman |
|---|---|---|---|---|
| | Rents | Net Flow Rates | Deviation | Rank Correlation |
| Earnings | | | | |
| 1970 | 0.78*** | 0.08 | 9.63 | |
| 1980 | 0.58*** | −0.19 | 6.99 | |
| 1990 | 0.93*** | 0.06 | 10.00 | |
| 2000 | 0.80*** | −0.27 | 7.98 | |
| | | | | 0.77*** |
| | Correlation with | | Standard | Spearman |
| | Earnings | Net Flow Rates | Deviation | Rank Correlation |
| Rents | | | | |
| 1970 | 0.78*** | 0.09 | 13.65 | |
| 1980 | 0.58*** | −0.27 | 11.17 | |
| 1990 | 0.93*** | 0.11 | 21.39 | |
| 2000 | 0.80*** | −0.31** | 15.57 | |
| | | | | 0.76*** |

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

A few observations are in order. First, the positive cross-sectional correlation between rents and earnings suggests that labor demand, as opposed to labor supply, shocks are the main sources of employment variation at the state level (Roback (1982)). Figure 4 reports a scatter plot of rents and earnings across U.S. states using the 2000 Census data. Second, 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, suggesting considerable persistence in state rankings. Last, the levels of both wages and rents in each Census year tend to display a correlation of approximately zero with net flow rates (see Figures 5 and 6).[9]

## 2.3 Accounting for Observable Differences Across States

In this section I briefly discuss whether differences in observables across states can account for some of the stylized facts reported in the previous section. The analysis reveals that the answer to this question is negative.

First, U.S. states display considerable variation in size and other attributes. This variation might have an impact on measures of gross and net flows of workers. For example, it could be argued that larger states might have lower outflow rates. To evaluate the importance of some of these factors, I ran a cross-sectional regression of inflow, outflow and net flow rates on states' land area, a variable that equals the number of neighboring states of each state, and dummies for whether the state borders with Mexico or Canada. I have then defined the adjusted outflow, inflow and net flow rates to be the residuals of this regression. These adjusted series are highly correlated with the original ones and all the stylized facts of the previous section apply to these adjusted rates as well.[10]

Second, U.S. states also have a different composition of population, in terms of age, education, industry of employment, etc. Thus, when interpreting the statistics about population flows reported in the previous sections, it is important to take possible composition effects into account. For example, certain states might exhibit higher gross flows because of the sectoral or demographic composition of their employment structure. To address this issue, I divided the population into 385 demographic groups defined by age, education, and industry. Then, I computed gross and net flow rates for each state and for each demographic group. Last, I computed the state-wide rates as a weighted average of the groups' rates using as a weight for each group its relative size in the U.S. population. It turns out that the flow rates obtained using this procedure are very close to the unadjusted ones.[11] Thus, composition effects due to cross-state heterogeneity in the age, education and industry affiliation of the states' population do not seem to play an important role in explaining differential gross population flows across states.

[9] The only exception to this pattern is represented by the negative statistically significant - but small - correlation between net flow rates and rents in the 2000 Census.

[10] The correlation coefficient between raw and adjusted rates is 0.92 for net flows, 0.94 for gross inflows and 0.99 for gross outflows. The regression has a small impact on the cross-sectional standard deviation of the three series.

[11] The cross-sectional correlation between adjusted and unadjusted rates in the 2000 Census is always above 0.97.

Among this last set of results, it is particularly surprising that a state's net flow of workers appears to be independent of its sectoral composition of employment. In order to further elaborate on this point, I have used yearly data on states' employment by one digit industry from the Bureau of Labor Statistics for the period 1969-2001. I have then compared a state's employment growth series between 1969 and 2001 with an adjusted series that controls for differences in the composition of employment across states. Specifically, the adjusted series is obtained by summing, for each state, the growth of employment by industry weighting each industry by its share of U.S. aggregate employment instead of its share of the state's employment. For all states the original employment growth series and the adjusted series display a correlation of about 0.99. Given that net migration accounts for most of the differences in states' relative employment growth, this result provides further support for the fact that differences in the composition of employment by industry do not play an important role in explaining the stylized facts of the previous section.

Third, the cross-sectional correlations of Table 2, particularly the one between gross inflows and gross outflows, are consistent with two alternative stories. According to the first one, these positively correlated gross flows are symptoms of a changing industry/demographic mix of states' workforce (Sjaastad (1962)). The second interpretation, instead, is that this positive correlation is driven by matching effects between locations and observationally homogeneous workers (Kennan and Walker (2005)). The former interpretation of the data would suggest that outgoing workers are observationally different from incoming ones. The latter interpretation would, instead, imply that gross flows occur mainly within narrowly defined demographic and industry groups. To distinguish between these two alternatives, I have used the 385 demographic groups (indexed by $g$ and described above) and computed, for each state $j$ and for different Census years, the following measure of within demographic group workers' reallocation:[12]

$$\frac{\sum_g (in_{jg} + out_{jg}) - \sum_g |in_{jg} - out_{jg}|}{\sum_g (in_{jg} + out_{jg}) - \left|\sum_g (in_{jg} - out_{jg})\right|}. \tag{1}$$

Notice that this index takes values between zero and one: a value of 1 occurs if all gross flows occur within groups ($in_{jg} = out_{jg}$ for all $g$) and value of zero occurs if all flows occur across groups (in this case, for each group $g$, we would either have $in_{jg} > 0$ and $out_{jg} = 0$ or $in_{jg} = 0$ and $out_{jg} > 0$.). The population-weighted average of this measure across states for the 2000 Census was 0.92, suggesting that most flows occur within the demographic/industry groups described above. Similar results obtained for the other Census years.

A way to consider exclusively within-group flows when computing the correlations of Table 2, is to compute the cross-sectional correlation between gross inflows and outflows separately for each group. Then, the 385 correlation coefficients can be averaged using as weights the groups' population shares in the U.S. The following Table reports these adjusted

[12] This measure has been used, for example, by Davis and Haltiwanger (1992) to decompose aggregate excess reallocation of jobs into a between-sector and a within-sector component.

correlations for the 2000 Census:[13]

Table 4
Adjusted Cross-Sectional Correlations (Census 2000)

| | Outflow Rate | Inflow Rate | Net Flow Rate |
|---|---|---|---|
| Outflow Rate | 1 | 0.42*** | −0.19 |
| Inflow Rate | | 1 | 0.79*** |
| Net Flow Rate | | | 1 |

Note: *** denotes significant at 1% level.[14]

Comparing Tables 2 and 4, one notices that the correlation pattern highlighted in stylized fact #3 is robust, although the correlations in Table 4 are a bit smaller.

# 3 Model

Given the discussion of the previous section, in the following subsections I construct a general equilibrium model of migration in which both workers and locations are ex-ante homogeneous.

## 3.1 Description

The model presented in this section 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 market in the model is a persistent shock to the total factor productivity of its inputs. 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. 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 a 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.[15]

[13] The positive correlation between gross inflows and outflows is a well-known, though not extensively documented, stylized fact in the literature on internal migration of population. In his review paper Greenwood (1975) writes that "the observation has frequently been made that where inmigration is great, outmigration is also great." Miller (1967) shows that this correlation holds also within demographic groups defined by race, sex and occupation (see her Table 3).

[14] Notice that each of the correlations reported in Table 4 is an average of 385 correlations computed using the 385 demographic groups. The significance level of each average correlation reported in Table 4 is set equal to the significance level of the median of the distribution of these 385 correlation coefficients.

[15] 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

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 a shock to the productivity of the production inputs. Firms solve static optimization problems hiring efficiency units of labor at the rate $w$ and units of land at the rate $r$ after observing the productivity shock $z$.

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

The shock $z$ plays a key role in the analysis so its interpretation deserves some comments. In Section 2, I argued that the observed population flows do not appear to be "explained" by differences in the sectoral shares of states' output. Thus, from this perspective, $z$ should not be interpreted as a purely sectoral shock that has the same effect on a given sector in all locations. Instead, $z$ should be interpreted as a factor that makes a specific location particularly attractive for firms in general, and also, possibly, for firms in specific sectors of the economy. 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. 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.[16] The introduction and diffusion of air conditioning in the South-West of the U.S. has also improved working conditions, in addition to individuals' quality of life. These kinds of considerations are captured, albeit in a reduced form way, by the local amenity $z$. According to this interpretation, temporal shifts in productive amenities have led firms to move to more favorable locations, inducing workers attracted by relatively high wages to follow them. Alternatively, Appendix C shows that the model can be reinterpreted as one in which the shock $z$ represents an amenity shock, in the sense that it affects workers' utility in a location while changing stochastically over time. Such model would have exactly the same implications for migration flows across states as the model considered in this section.[17]

**Preferences and workers' migration**. The economy is populated by a measure one of ex-ante identical and infinitely-lived dynasties. Each dynasty is characterized by the following

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.

[16] See Peet (1983) for a discussion of the connection between changes in the distribution of the manufacturing industry across U.S. states and states' rates of unionization. Lee (2004) and the references therein contain evidence about the effect of states' fiscal policies on the geographic distribution of economic activity.

[17] Notice that the positive cross-sectional correlations between state-level earnings and rents (Figure 4) suggest that the driving force behind internal migration flows are labor demand shocks. This is consistent with previous work by Blanchard and Katz (1992) and Topel (1986).

preferences:[18]

$$E\left[\sum_{t=0}^{\infty}\beta^{t}\left(c_{t}+\phi\left(l_{t}^{c}\right)-\upsilon_{t}\right)\right],$$

where $c_t$ denotes consumption of goods, $l_t^c$ represents consumption of land, and $\upsilon_t$ is a location-specific shock to utility. The function $\phi$ is such that $\phi' > 0$, $\phi'' < 0$, and $\lim_{x\to 0}\phi'(x) \to \infty$. The parameter $\beta < 1$ represents the discount factor. At a point in time only one worker is alive in each dynasty. This worker faces a probability $1-\delta$ of dying in each period. If he dies, he is replaced by another worker of the same dynasty who begins his life in a location chosen randomly. Upon birth in a location $j$, an agent $i$ draws the utility shock $\upsilon_{ijt}$. The latter represents an agent's match with the location and remains the same as long as the agent stays in the same location. An agent can choose to leave a location and relocate to a different one at a cost $k$. By moving to a location $j'$ at the end of period $t$, the agent will draw another location-specific utility shock $\upsilon_{ij't+1}$.

An agent $i$'s flow budget constraint in period $t$ is:[19]

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

In detail, the sequence of events in an agent's life is as follows:

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

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

- The agent optimally chooses $c$ and $l^c$ and receives a utility flow $(c + \phi(l^c) - \upsilon)$.
- With probability $1-\delta$ the agent dies and is replaced by another agent of the same dynasty that starts his life in a random location at the end of period $t$. In this random assignment, the probability that a newly born agent starts his 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. In this case, he can then 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. He if decides to move he obtains expected utility $e$.

[18]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.

[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.

- At the beginning of period $t+1$, if the agent had remained in the same location in which he was living in $t$, he keeps the same location-specific utility shock $\upsilon$. If the agent has chosen to move to a new location, he draws a new idiosyncratic location-match from the distribution (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 $\eta$ 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 himself, 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 $e_r$. As for newly born individuals, also in this case the probability that an agent is reallocated 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:

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

**Value Function.** At the beginning of a period the state of a location is fully characterized by the vector $(y, n, \zeta)$, where $n$ denotes the measure of workers with location-specific match shock $\upsilon_1$. Denote the laws of motion for $y$ and $n$ by $Y(s,\zeta)$ and $N(s,\zeta)$ respectively, and summarize the two in the law of motion for $s=(y,n): s' = S(s,\zeta)$, where $S(s,\zeta) = [Y(s,\zeta), N(s,\zeta)]$. The value function of a worker characterized by idiosyncratic match $\upsilon$ with the location is given by:

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

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

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

where the first term on the right-hand side of this equation represents the endogenous component (due to directed search) of gross inflows into a location while the second term represents

[20] A natural interpretation of "undirected search" in models of geographical mobility is that an agent is drawn to a given location by exogenous idiosyncratic factors that are independent of the state of that economy.

the exogenous one (due to undirected search). The latter is equal to the product of the aggregate level of inflows $\overline{x}$ and the share $y\left(1-\eta\right)$ of inflows that are due to random reallocation of workers.

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

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

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

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

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

$$\mu\left(S',\Xi'\right)=\int_{\{s,\zeta:s'\in S'\}}Q(\zeta,\Xi')\mu\left(ds,d\zeta\right). \tag{8}$$

## 3.2 Equilibrium

A recursive stationary equilibrium for this economy is represented by a value function $V\left(s,\zeta,\upsilon\right)$, a probability $q\left(s,\zeta,\upsilon\right)$, laws of motion $S\left(s,\zeta\right)=\left[Y\left(s,\zeta\right),N\left(s,\zeta\right)\right]$, 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 $\mu\left(s,\zeta\right)$, demand for land by consumers $l^c\left(s,\zeta\right)$, demand for land by firms $l^f\left(s,\zeta\right)$, a wage function $w(s,\zeta)$, and a rent function $r(s,\zeta)$ such that:

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

$$r(s,\zeta)=\phi'\left(l^c(s,\zeta)\right). \tag{9}$$

- The law of motion $S\left(s,\zeta\right)$ is related to $x(s,\zeta)$, $o(s,\zeta)$ and $q\left(s,\zeta,\upsilon\right)$ by equations (6) and (7).

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

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

$$e_d\geq p\int V\left(S\left(s,\zeta\right),\zeta',\upsilon_l\right)Q(\zeta,d\zeta')+(1-p)\int V\left(S\left(s,\zeta\right),\zeta',\upsilon_h\right)Q(\zeta,d\zeta').$$

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

$$e_d=p\int V\left(S\left(s,\zeta\right),\zeta',\upsilon_l\right)Q(\zeta,d\zeta')+(1-p)\int V\left(S\left(s,\zeta\right),\zeta',\upsilon_h\right)Q(\zeta,d\zeta').$$

- 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:

$$e_r = p\int V\left(s,\zeta,\upsilon_l\right)y\mu\left(ds,d\zeta\right) + (1-p)\int V\left(s,\zeta,\upsilon_h\right)y\mu\left(ds,d\zeta\right).$$

- Aggregate population has measure one:

$$\int y\mu\left(ds,d\zeta\right) = 1.$$

- The invariant distribution $\mu(s, \zeta)$ is consistent with individual decisions, so equation (8) holds.
- The wage and rent functions in each location are such that:

$$\begin{aligned} r(s,\zeta) &= \zeta F_l\left(y, l^f(s,\zeta)\right), \\ w(s,\zeta) &= \zeta F_y\left(y, l^f(s,\zeta)\right). \end{aligned} \tag{10}$$

- The land market in each location clears:

$$yl^c(s,\zeta) + l^f(s,\zeta) = 1. \tag{11}$$

- Aggregate gross inflows are given by:

$$\overline{x} = \int x(s,\zeta)\mu\left(ds,d\zeta\right).$$

# 4 Empirical Implementation

## 4.1 Estimation

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 constant values of migration $e_d$ and $e_r$. The empirical moments were, instead, constructed using data from the 48 contiguous U.S. states. 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 full distribution of the state vector $(s, \zeta)$ across all other locations when solving his dynamic programming problem. Notice that, in the data presented in Section 2.2, the cross-sectional properties of both gross and net flows tend to be quite similar across Census decades, which is consistent with the equilibrium of the model with a continuum of locations.

Some of the model's parameters are set a-priori and others are estimated using 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 $\beta$ is set equal to 0.82, implying a yearly interest rate of 4 percent. An individual's working life in the data I consider is about 30 years, or 6 model-periods. Thus, I set the constant probability of survival $\delta$ equal to 0.83, so that the average lifetime for an individual in the model is approximately 6 periods.

The moving cost parameter $k$ is set equal to 0.1. This represents 11.5% of the average wage income in the model. To place this value in perspective, notice that the average wage 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\$. In section 4.3 I conduct some sensitivity analysis with respect to the magnitude of this parameter.

The random variable $\upsilon$ can take one of two values, $\upsilon_1$ and $\upsilon_2$, with probability $p$ and $1-p$, respectively. The shock $\upsilon_2$ is normalized to zero. The choice of $\upsilon_1$ is important for the properties of the model. The value of this parameter that I pick reflects a balance between trying to account for the properties of gross flows observed in the data and keeping the model manageable. The basic trade-off is that, on the one hand, to keep the model from becoming exceedingly complicated, it is necessary to choose a value of $\upsilon_1$ that implies that poorly matched workers always migrate independently of the state of the local economy where they reside. On the other hand, if $\upsilon_1$ is too large, agents who draw a high match with a location will never choose to leave because of the potential loss of their idiosyncratic match. If this were the case, the model could not account for the U-shaped relationship between net flow rates and gross outflow rates in the cross-section. The value of $\upsilon_1$ that I choose is 0.08. This number represents the *minimum* value of $\upsilon_1$ such that poorly matched workers always choose to migrate (i.e., $q(s, \zeta, \upsilon_1) = 1$). Appendix B provides details on the characterization of the model's equilibrium in a situation in which poorly matched workers always choose to migrate.

The parameter $\eta$ 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 ("indirected search"). I set $\eta$ on the basis of the observation that even the U.S. states characterized by negative net flow rates tend to attract relatively large flows of workers in gross terms. Specifically, notice from equation (4) that the gross inflow rate into a location for which voluntary inflows ($x_d(s,\zeta)$) are zero is $\delta^{-1}\overline{x}\,(1-\eta)$, where $\overline{x}$ denotes the average inflow rate for the entire economy. I set $\eta$ to equalize this rate in the model with the observed gross inflow rate for North Dakota in the 2000 Census (7.39%). North Dakota is the U.S. state with the lowest net flow rate in the 2000 Census sample. Given the calibrated value of $\delta$ and the fact that the model can by construction reproduce the observed $\overline{x}$ (the parameter $p$ is set to match that moment - see below), this procedure yields $\eta = 0.2876$.

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

$$F(m,l) = m^{\tau} l^{1-\tau},\ \tau \in (0,1)\,, \tag{12}$$

$$\phi\,(l) = \frac{A}{\alpha} l^{\alpha},\ \alpha \in (0,1)\,,\ A > 0. \tag{13}$$

The parameter $\tau$ 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 $\tau$ 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 (0.06 and 0.60 respectively). The elasticity parameter $\alpha$ 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.

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

$$z' = z\varepsilon', \tag{14}$$

$$\varepsilon' = z^{\psi-1}\varepsilon^{\rho}u' \tag{15}$$

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 $\left(\exp\left\{\sigma_u^2\right\}-1\right)$. The specification of the exogenous shocks in equations (14)-(15) is non-standard because it assumes that the growth rate of productivity $\varepsilon'$ in (14) is persistent, as opposed to being identically distributed over time. Persistence in the process followed by $\varepsilon'$ is necessary in order to generate persistent net flows in the Lucas-Prescott model. If $\rho$ were equal to zero, net flows would be negatively autocorrelated over time, which is strongly at odds with the data (see Table 3).[21] In order to obtain a stationary process for $\{z\}$ it is necessary that the parameter $\psi$ be strictly smaller than one. Ideally both parameters $\rho$ and $\psi$ should be identified by the autocorrelation structure of states' net flow rates, with lower

[21] 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.

persistence in the latter being associated with smaller values of $\rho$ and $\psi$. In practice, the limited panel dimension for states net flow rates (see Table 3) only allows identification of the parameter $\rho$. The value of the parameter $\psi$ 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 $\theta = (\rho, \sigma_u, p, A)$. First, $\rho$ and $\sigma_u$ 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 $\sigma_u$ is identified by the cross-sectional standard deviation of net inflow rates in the 2000 Census (0.0286 - from Table 1, 2000 Census) and the parameter $\rho$ by the first-order autocorrelation coefficient of net inflow rates across Census years (0.58 - from Table 3).

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 observed interstate migration rate in the U.S. economy, in the 1995-2000 period. From Table 1, this value is 8.61 percent.

Third, the parameter $A$ is set to match the observed dispersion of residual average wage income across U.S. states in the 2000 Census. From Table 4, this measure of dispersion (the standard deviation of the log of wage income) is 7.98. 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.

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

Table 5
The Model's Parameters

| | Parameter | How it is set | Estimate |
|---|---|---|---|
| $\beta$ | discount factor | a-priori | 0.8200 |
| $\delta$ | probability of death | a-priori | 0.8300 |
| $\tau$ | production function parameter | a-priori | 0.9091 |
| $\alpha$ | utility function parameter | a-priori | 0.0909 |
| $\upsilon_1$ | value of low idiosyncratic shock | a-priori | 0.0800 |
| $\upsilon_2$ | value of high idiosyncratic shock | a-priori | 0.0000 |
| $\eta$ | search parameter | a-priori | 0.2876 |
| $\psi$ | mean-reversion parameter for $z$ | a-priori | 0.9990 |
| $\rho$ | first order autocorrelation of $\varepsilon$ | estimated | 0.7868 |
| $\sigma_u$ | volatility of labor demand shock | estimated | 0.0031 |
| $p$ | probability of low idiosyncratic shock | estimated | 0.3521 |
| $A$ | utility function parameter | estimated | 0.0900 |

Two features of the parameter estimates stand out. First, the growth rate of productivity is found to exhibit very large persistence with a first-order autocorrelation coefficient of about 0.78 at 5-year intervals. Second, the estimate of the parameter $A$ implies that a worker spends, on average, 12% of his labor income on the fixed factor. This is reasonable. According to Roback (1982) the budget share of land is around 6%, but recent estimates (e.g. Heathcote and Morris (2007)) of the value of land as a share of the total value of residential real estate suggest a larger figure.

## 4.2 Results and Intuition

The following tables represent the main cross-sectional and time-series statistics generated by the model. I emphasize in bold the moments that were targeted in the estimation of the model. For convenience, the tables also report the corresponding data moments for the 2000 Census.

Table 6
Migration Flows

| | Mean | | St. Deviation | | Corr. Outflow | | Corr. Inflow | |
|---|---|---|---|---|---|---|---|---|
| | Data | Model | Data | Model | Data | Model | Data | Model |
| Outflow Rate | **8.61** | **8.61** | 1.72 | 0.93 | 1 | 1 | 0.69 | 0.56 |
| Inflow Rate | 8.61 | 8.61 | 3.75 | 3.28 | 0.69 | 0.56 | 1 | 1 |
| Net Flow Rate | 0.00 | 0.00 | **2.86** | **2.86** | 0.30 | 0.32 | 0.90 | 0.96 |

Note: The data columns refer to the 2000 Census.

Table 7
Autocorrelations of Migration Flows

| | Autocorrelation Coefficients | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $t, t-1$ | | $t, t-2$ | | $t, t-3$ | | $t, t-4$ | |
| | Data | Model | Data | Model | Data | Model | Data | Model |
| Outflow Rate | 0.76 | 0.52 | 0.79 | 0.27 | 0.67 | 0.13 | 0.60 | 0.05 |
| Inflow Rate | 0.88 | 0.67 | 0.83 | 0.48 | 0.84 | 0.35 | 0.68 | 0.27 |
| Net Flow Rate | **0.58** | **0.58** | 0.61 | 0.42 | 0.53 | 0.32 | 0.33 | 0.24 |

Note: The data columns refer to the 2000 Census.

Table 8
Wages, Rents and Net Flows

| | Corr. Net Flow Rate | | Corr. Rent | | Dispersion | | Rank Corr. | |
|---|---|---|---|---|---|---|---|---|
| | Data | Model | Data | Model | Data | Model | Data | Model |
| Earnings | $-0.27$ | $0.17^{***}$ | $0.80^{***}$ | $0.96^{***}$ | **7.98** | **7.98** | $0.77^{***}$ | $0.99^{***}$ |
| Rent | $-0.31^{**}$ | 0.02 | 1 | 1 | 15.57 | 83.42 | $0.76^{***}$ | $0.99^{***}$ |

Note: The data columns refer to the 2000 Census. $^{***}$ significant at 1% level. $^{**}$ significant at 5% level.

The simulated model is generally consistent with the stylized facts presented in Section 2. Specifically:

**Stylized fact #1 (Table 6).** By construction, the model accounts for the fact that gross flows are large relative to net flows. There are two main determinants of gross flows in the model. First, in some locations, declining productivity induces some highly matched workers to leave to search for locations with better aggregate characteristics. Second, some workers have a poor match with their location of residence and decide to move for idiosyncratic reasons. In the model, out of every ten moves, nine occur for the latter reason, while one occurs for the former one. Thus, idiosyncratic shocks are the primary source of labor mobility across states. Consistently with the evidence (see e.g. Greenwood (1975)), the model predicts that migration rates are higher for younger workers. This is due to a selection effect: a younger worker has had less time to find a location with a high value of the idiosyncratic match shock $\upsilon$. Kennan and Walker (2006) point out that in the NLSY data 53 percent of movers moves more than once, which is inconsistent with a simple binomial model in which all workers face the same probability of moving in every period. In this model, the proportion of movers who move more than once in three model-periods (the rough equivalent of 13 years in Kennan and Walker's sample) is 35 percent, while a simple binomial model would yield a much lower proportion of 8 percent. Again, selection effects account for the uneven distribution of moves in the population.

**Stylized fact #2 (Table 6 and Figure 7).** In the model's equilibrium, gross inflow rates are cross-sectionally more dispersed than net flow rates, while the latter are more dispersed than gross outflow rates. It is interesting to notice that in the model gross inflow rates display the highest dispersion across locations, despite the fact that only less than a third of all moves is directed toward a location with good aggregate state. In the limit, if $\eta = 0$ (completely undirected search), all locations would display the same inflow rate. The cross-sectional dispersion of inflow rates in the model has to do with the fact temporal variations in gross inflows are a more efficient way of adjusting a location's employment level in response to location-specific productivity shocks than temporal variations in gross outflows. For example, to generate higher gross outflows, as opposed to lower gross inflows, in response to a negative productivity shock, workers with a high idiosyncratic match with the location would have to migrate. The loss of this idiosyncratic match component represents an implicit cost of adjusting gross outflows. Of course, when the location's aggregate productivity display a relatively large decline, adjustment of gross inflows becomes unfeasible (i.e., $x_d(s, \zeta) = 0$ in my notation), and gross outflows have to increase. Figure 7 represents cross-sectional scatter plots of gross inflow and outflow rates against net flow rates. The data have been generated by the benchmark version of the model. The next point discusses further the interpretation of this figure.

**Stylized fact #3 (Table 6 and Figure 7).** The cross-sectional correlation patterns among workers flows are consistent with the data. The most interesting features of the data, which the model can account for, are the positive correlation between gross inflow and outflow rates and the U-shaped relationship between net flows and gross outflows.

Consider the former first. As shown in Figure 7, locations characterized by higher gross inflow rates also tend to display relatively large outflow rates. This is due to the selection effect already mentioned above because incoming workers tend to have a lower average match with their new location than incumbent ones. Thus, a large inflow rate in a location is associated with a subsequent large outflow rate from the same location. The U-shaped relationship between outflow rates and net flow rates reflects two different forces that are evident in Figure 7. On the one-hand, for values of net flow rates above $-1$ percent, variations in gross inflows represent the main mechanism of adjustment of a location to productivity shocks. Due to the selection effect, in this region large gross inflow rates are associated with large gross outflow rates. On the other hand, for values of net flow rates below $-1$ percent, gross inflow rates reach their lower bound (i.e., $x_d(s, \zeta) = 0$), and the location adjusts to negative productivity shocks by means of higher gross outflows. Hence, gross outflow rates are relatively high for both high and positive and high and negative net flow rates.

**Stylized fact #4 (Table 7).** All workers' flows in the model are persistent over time, even if somewhat less than in the data. The key to persistence of all flows is the fact that the growth rate of productivity, $\varepsilon$, is itself persistent ($\rho > 0$). If the productivity shock $z$, instead of its growth rate $\varepsilon$, 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.

**Stylized fact #5 (Table 8 and Figures 8-10).** Both in the model and in the data, the cross-sectional correlation between net flow rates and average earnings and rents is very small (Figures 8 and 9). The intuition for this result is that net flow rates depend on the *growth rate* of total factor productivity in a location, $\varepsilon$, 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 (Figure 10). The model correctly predicts that rents are more dispersed than earnings in the cross-section, but it strongly overpredicts the magnitude of rent dispersion.

## 4.3 Sensitivity Analysis on the Size of Moving Costs

In a recent paper, Kennan and Walker (2006) estimate a partial equilibrium model of migration and obtain very large estimates of moving costs. For example, according to their Table 3, the moving cost for the average mover is between \$229,000 and \$176,000 (figures in 2005 \$), according to the exact specification of the model. I use this as a motivation to ask what effect a significantly higher moving cost than the one assumed in Section 4 would have on the aggregate moments of interest. Specifically, the experiment consists of increasing the cost of migration by a factor of ten, from \$25,000 to \$250,000, while simultaneously increasing the idiosyncratic disutility of a poor match with a location, captured by the parameter $\upsilon_1$, from 0.08 to 0.52. The latter is the minimum value of $\upsilon_1$ that guarantees that poorly matched workers always choose to migrate.[22] Given these new values for $k$ and $\upsilon_1$, I re-estimate the model along the lines described in Section 4. The parameters' estimates are very close to the original ones reported in Table 5. The implied properties of net and gross flows are reported in Table 9.

Table 9
Migration Flows in Model with Large Moving Costs

| | Mean | | St. Deviation | | Corr. Outflow | | Corr. Inflow | |
|---|---|---|---|---|---|---|---|---|
| | Data | Model | Data | Model | Data | Model | Data | Model |
| Outflow Rate | **8.61** | **8.61** | 1.72 | 0.81 | 1 | 1 | 0.69 | 0.82 |
| Inflow Rate | 8.61 | 8.61 | 3.75 | 3.49 | 0.69 | 0.82 | 1 | 1 |
| Net Flow Rate | 0.00 | 0.00 | **2.86** | **2.86** | 0.30 | 0.72 | 0.90 | 0.98 |

Note: The data columns refer to the 2000 Census.

As the Table shows, the main difference between the results obtained in the economy with high moving costs and the results obtained in the benchmark specification of the model concerns the correlation coefficients between the outflow rate on the one hand and net and gross inflow rates on the other. These correlations are significantly higher in the model with high moving costs. The intuition is simple. In response to negative productivity shocks, a location can adjust by either attracting fewer workers from the rest of the economy, or by experiencing higher rates of outmigration because highly matched workers choose to leave. In this version of the model, this type of workers never migrates because of the high moving costs. Therefore, when moving costs are high, the U-shaped relationship between net flow rates and outflow rates documented in Section 2.2 is lost. Outflow rates still tend to be large for locations experiencing relatively high gross and net inflows. This is due to the selection effect discussed above: poorly matched workers are a disproportionate fraction of the population in these locations. However, due to the large moving costs, outflow rates are low in location experiencing net losses of population. The other moments implied the model with large moving costs are quite similar to the ones reported in Tables 6-8 above and are therefore not reported.

In summary, the fit of the model with lower moving costs is somewhat better than the one with large moving costs. While parameter estimates are always a function of a model's specification and focus, I think that the intuition obtained here is, at least qualitatively, applicable to Kennan and Walker's model as well, given the similar focus on idiosyncratic match shocks as primary motives from migration in both papers. The tension identified in this paper is the following. On the one hand, Kennan and Walker point out that large moving costs are needed to account for the fact that inter-state migration rates are relatively small, despite potentially large gains from geographical mobility. As in my model, these gains have both an idiosyncratic match components and an location-specific one. On the other hand, a model with idiosyncratic match shocks as well as aggregate shocks, such as mine, predicts that the type of workers who are left in a location with declining relative population are the workers with better idiosyncratic matches with the location. The other workers are the first ones to leave. Now, these highly matched workers would not want to migrate for purely idiosyncratic reasons because, if they did, they would draw a worse idiosyncratic match in

[22] If this parameter stayed constant, nobody would ever migrate in this version of the model.

expectation. Therefore, they would only want to move to take advantage of higher average wages elsewhere in the economy. However, state-level differences in wages and rents alone might not be large enough to overcome the disincentives of large moving costs for this group of workers. The calibrated version of my model suggests that this is indeed the case.

# 5 Discussion and Implications: Gross Flows vs. Net Flows

In an economy like the one considered in this paper, in which gross and net flows do not coincide, a location might adjust to a downturn in local economic conditions by either absorbing fewer workers from the rest of the economy or by experiencing higher outmigration. In this section I argue that these two modes of adjustment have, in principle, different observable implications for wages and rents.

In this model economy outmigration is more costly, as an adjustment mechanism, than reduced immigration because it entails two additional costs: 1. a direct moving cost for workers who move out of a location; 2. an indirect moving cost, as these workers lose their high idiosyncratic match with the location in exchange for a smaller expected match with other locations. In principle, instead, if the local economy could adjust by absorbing fewer immigrants from the rest of the economy, it would avoid both type of costs. Of course, when gross inflows reach their lower bound, only one channel of adjustment is left, i.e. outmigration.[23] The intuition for this is that there is a group of workers who always migrate from any location for idiosyncratic reasons. This "natural" decline in the workforce of each location allows variations in gross inflows to play a role as locations who experience negative productivity shocks can adjust by simply avoiding to replace the workers who left for idiosyncratic reasons. The closest analogy to this is that of a corporation in which a given number of workers retires at the end of every year. If the company has to reduce employment it might achieve this goal simply by hiring fewer new workers to replace the retiring ones, rather than "firing" any current worker. Of course, if the required reduction in employment is large enough, "firing" (i.e., outmigration) becomes a necessary option.

An implication of this discussion is that the exact way in which the local economy adjusts to shocks - through smaller inflows or higher outflows - will be reflected in the dynamics of local prices such as wages and rents. To see this, consider the case in which the shock is large enough that higher outflows are required to bring the local economy back to equilibrium. Migrating workers bear the direct and indirect costs of adjustment mentioned above. Observationally equivalent workers who do not migrate do not bear these costs, but must experience the same drop in expected utility as migrating workers. This drop is achieved by a reduction in local wages. When, the shock is accommodated through lower inflows, instead, local wages do not necessarily have to fall.

This implication of the model can, in principle, be tested empirically by checking whether the correlation between percentage changes in earnings and net flow rates is higher for location in which adjustment takes the form of lower inflows or, instead, takes the form of

[23] The model also displays an area of inaction in which no highly matched worker leaves the location and no potential migrant chooses to settle there. The results below take this case into account.

higher outflows. In practice, in order to identify these regions in a way that is consistent with the theoretical model is not straightforward. For example, one would need to be able to assign each state-year observation in the Census data to one of the three regions that characterize the equilibrium of the benchmark model (see Appendix B). These regions are as follows. Region 1: Some outflows by highly matched workers, no voluntary inflow of workers from the rest of the economy. Region 2: No outflows by highly matched workers, no voluntary inflow of workers from the rest of the economy. Region 3: No outflows by highly matched workers, some voluntary inflow of workers from the rest of the economy.

To do so, I follow this procedure:

**Step 1.** Simulate data from the benchmark version of the model. Notice that in Regions 1 and 2, gross inflows are relatively small because the location does not attract any voluntary migration from the rest of the economy. Thus, Regions 1 and 2, which account for about 40 percent of the entire observations, are easily identifiable in the Census data. In practice, I do so by ranking the state-year observations in the pooled 1970-2000 Census data by their gross inflow rate and by assuming that a state-year observation belongs to either Region 1 or 2 if it belongs to the bottom 40 percent of this distribution. A state-year observation is, instead, assumed to belong to Region 3 if its gross inflow rate is in the top 60 percent of this distribution.

**Step 2.** To distinguish between Region 1 and Region 2, notice that in Region 1 outflow rates tend to be larger than in Region 2 because in the former highly matched workers tend to migrate while in the latter they do not. Based on this intuition, I use the observations generated by the benchmark version of the model that belong to Regions 1 and 2 to estimate a logit model in which the explanatory variable is the outflow rate. The logit provides a way to map migration rates, which can be observed in both model-generated data and Census data, into the probability that an observation belongs to either Region 1 or 2. The estimated logit then allows me to assign to each year-state observation $(T, j)$ in the Census data a probability, denoted by $P_{Tj}^{r}$ $(r = 1, 2)$, that this observation belongs to Region $r = 1, 2$.[24] The results are summarized in Table 10.

Table 10
Estimated Average Probabilities of Census State-Year Observations
Belonging to Regions 1, 2, 3

| | Probability of State-Year Obs Being in | | |
|---|---|---|---|
| Observed Net Flow Rate | Region 1 | Region 2 | Region 3 |
| Positive (97 obs) | 0% | 22% | 78% |
| Negative (95 obs) | 30% | 28% | 42% |
| $< -5\%$ (12 obs) | 67% | 0% | 33% |
| $> -5\%$ (83 obs) | 25% | 32% | 43% |

[24]Recall from Step 1 that $P_1 = P_2 = 0$ and $P_3 = 1$ if gross inflows exceed the 40-th percentile of the distribution. If gross inflows are below the 40-th percentile the assignment of probabilities is based on the logit.

The table gives the average (across observations) probability that a state-year observation in the Census belongs to each of the three Regions, conditional on its associated net flow rate being positive or negative. The results indicate that observations with positive net flow rates are very likely to come from local economies in Region 3, while observations with negative net flow rates might come from each of the three Regions, with slightly higher chances of being in Region 3. On the other hand, locations experiencing large negative net flow rates are much more likely to be in Region 1. This is reasonable since for those locations gross inflows are more likely to be at their lower bound.

**Step 3.** Last, I use these probabilities to run the following regression:

$$\log\left(\text{Earnings}_{Tj}\right) - \log\left(\text{Earnings}_{Tj-10}\right) = \text{Census year dummies} + \text{Net flow rate}_{Tj} \times \sum_{r=1}^{3} \psi^r \times P^r_{Tj} + e_{Tj}. \quad (16)$$

In this regression, $T = 2000, 1990, 1980$ denotes Census years and $j = 1, 2, .., 48$ a U.S. state. "$\text{Earnings}_{Tj}$" refers to state $j$'s fixed effect obtained from a cross-sectional regression in which individual weekly earnings from the year $T$'s Census are regressed on workers' observable characteristics (see Appendix A). Census year dummies control for economy-wide changes in the level of prices and other sources of variations in real earnings between a Census year and the previous one. The regressor "$\text{Net flow rate}_{Tj}$" represents the net flow rate into location $j$ between year $T-5$ and the Census year $T$. The parameters of interest are the $\psi^r$, $r = 1, 2, 3$. $\psi^r$ represents the elasticity of state-level earnings to a one percentage point increase in the net flow rate in a state that is estimated to be in Region $r$ with probability one in year $T-5$.[25] In other words, the coefficients $\psi^r$ provide information about how the correlations between changes in state-level earnings and relative population growth vary with the Region in which the local economy is presumed to be.

The argument advanced above suggests that we should expect $\psi^1 > \psi^3 \geq 0$. That is, we should expect that in response to a shock that makes net flows into a state negative, local earnings should drop more when the economy is in Region 1 than if the economy is in Region 3. When the economy is in Region 2, instead, workers' flows do not respond to local shocks and wages should display the greatest correlation with net flow rates. Formally, we should expect $\psi^2 > \psi^1$.

[25] Notice that if the Region of a state-year observation in the Census data was known with probability one, this would be a standard regression in which the variable net flow rate is interacted with dummy variables indexing the three possible Regions. The regression (16) generalizes such dummy variable regression when the Census data can be assigned to the three Regions only in a probabilistic way.

Table 11 reports the estimated coefficients of interest, $\psi^r$, for $r = 1, 2, 3$, in equation (16).

Table 11
Regression of % Changes in State-Level Earnings
on Net Flow Rates, Conditional on Three Regions

| | $\widehat{\psi}_1$ | $\widehat{\psi}_2$ | $\widehat{\psi}_3$ |
|---|---|---|---|
| Estimated Value | $0.84^{***}$ | $2.07^{***}$ | $0.29^{*}$ |
| Adjusted $R^2$ | | 0.94 | |
| F-Test: $H_0 : \widehat{\psi}_1 = \widehat{\psi}_3$ | | $3.00^{*}$ | |
| Number of observations | | 143 | |

Note: $^{***}$ significant at 1% level, $^{**}$ significant at 5% level, $^{*}$ significant at 10% level

The values of the estimated coefficients are consistent with the implication of the structural model discussed above. Specifically, it is the case that: $\widehat{\psi}_2 > \widehat{\psi}_1 > \widehat{\psi}_3$ and the estimates $\widehat{\psi}_1$ and $\widehat{\psi}_3$ are statistically different from one another.

In summary, the structural model presented in this paper implies that negative net flows from a location can be accommodated in two ways: lower gross inflows or higher gross outflows. The model also implies that wages in a location should be more affected by a negative shock in the latter situation than in the former. Analysis of wage changes using Census data is in accordance with this prediction of the model. This is an important result because, as the figures in Table 10 suggest, variations in gross inflows are the most important single channel through which negative shocks to employment are accommodated. It follows that studies of the joint analysis of the behavior of local wages and employment (see e.g. Blanchard and Katz (1992)) that do not distinguish between employment drops due to lower gross inflows of workers from those due to higher gross outflows, might reach the wrong conclusions. For example, observing a decline in local employment without a corresponding decline in local wages might lead to the conclusion that the adjustment process is inefficient due to "rigid" wages. However, this pattern is actually consistent with the model if the economy is in Region 3 during the adjustment period.

# 6 Summary

This paper makes two contributions. First, it presents in a systematic way the main stylized facts about net and gross flows of workers across U.S. states. Then, it introduces and estimates a model of workers' flows across locations. The model is consistent with the main features of the data. In particular, it is able to account for the lower cross-sectional dispersion of gross outflow rates than both gross and net inflow rates. The latter observation points to the importance of gross inflows, rather than gross outflows, as channel through which the state economy adjusts to local shocks. The model embeds both kinds of adjustments. In simulations, temporal variations in gross inflows appear to be the standard channel of adjustment to shocks. The outflow channel is active intermittently for states experiencing large negative net flows.

This paper emphasizes the role played by persistent shocks to the growth rate of state-level total factor productivity in order to generate persistent differences in net flow rates across states. A productivity shock in this paper is a convenient "catch-all" variable meant to capture all the factors that make firms want to locate in a certain state. Its interpretation remains an open question that deserves further scrutiny to uncover the more primitive forces behind it. Section 2.3 presented some evidence that seems to rule out sectoral shocks as their main factor associated with these forces. Candidate explanations for the productivity shock include slow-changing local institutions (Berkowitz and Clay (2004)) and time-varying differences in state-level policies (Lee (2004) and Peet (1983)). Future research will have to discriminate between these and other alternatives.

## 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, pp. 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.

Barro, Robert and Xavier Sala-i-Martin (1991): "Convergence Across States and Regions," *Brooking Papers on Economic Activity* 1991, 107-158.

Barro, Robert and Xavier Sala-i-Martin (1992): "Convergence," *Journal of Political Economy* 100, 223-251.

Berkowitz, Daniel and Karen Clay (2004): "Initial Conditions, Institutional Dynamics and Economic Performance: Evidence from the American States," William Davidson Institute Working Paper Number 615.

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

Borjas, George (2005): "Native Internal Migration and the Labor Market Impact of Immigration," NBER Working Paper 11610.

Borjas, George, Freeman Richard and Lawrence Katz (1997): unpublished appendix to "How Much Do Immigration and Trade Affect Labor Market Outcomes?," *Brooking Papers on Economic Activity* 1, 1-67.

Borjas, George, Bronars Stephen and Stephen Trejo (1992a): "Self-Selection and Internal Migration in the United States," *Journal of Urban Economics* 32, 159-185.

Borjas, George, Bronars Stephen and Stephen Trejo (1992b): "Assimilation and the Earnings of Young Internal Migrants," *Review of Economics and Statistics* 74, 170-175.

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

Dahl, Gordon (2002): "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica* 70, 2367-2420.

Davis, Morris and Jonathan Heathcote (2007): "The Price and Quantity of Residential Land in the United States," *Journal of Monetary Economics*

Davis, Steven and John Haltiwanger (1992): "Gross Job Creation, Gross Job Destruction, and Employment Reallocation," *Quarterly Journal of Economics* 107, 819-863.

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

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.

Hassler, John, Rodriguez-Mora, Jose', Storesletten, Kjetil, and Fabrizio Zilibotti (2005): "A Positive Theory of Geographic Mobility and Social Insurance," *International Economic Review 46*, 263-303.

Kennan, John and James Walker (2005): "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 Yoonsoo (2004): "Geographic Redistribution of U.S. Manufacturing and the Role of State Development Policy," Working Paper 04-15, Federal Reserve Bank of Cleveland.

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.

McCall, John (1970): "Economics of Information and Job Search," *Quarterly Journal of Economics* 84, 113-126.

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.

Mincer (1978): "Family Migration Decisions," *Journal of Political Economy* 86, 749-773.

OECD (2000): *Employment Outlook*, Paris.

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.

Rappaport J. (2007): “Moving to Nice Weather,” *Regional Science and Urban Economics* 37, 375–398.

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 (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

### 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).[26] 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 (as of their last birthday);
- 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.

In what follows, I will refer to the selected sample as the "population". The number of selected individual observations is , representing  million people.

### Measures of 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 $\tau$. 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 $\tau$ 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 $\tau$, reported living in location $j$ in year $\tau - 5$. Total outflow of population from location $j$ between $\tau - 5$ and $\tau$ is then defined as

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

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

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

[26] 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. I exploit the 1950 Census year in Table 3.

Let $y_{j\tau}$ denote the population living in location $j$ in Census year $\tau$:

$$y_{j\tau} = \sum_i \mu_{i\tau} I_{i\tau}(j).$$

I also denote by $\overline{y}_{j\tau}$ the total population that was interviewed in year $\tau$'s Census 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 $\tau$ is then defined as follows:

$$\widehat{o}_{j\tau} = \frac{out_{j\tau}}{\overline{y}_{j\tau}}.$$

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

$$\widehat{x}_{j\tau} = \frac{in_{j\tau}}{\overline{y}_{j\tau}}.$$

The net flow *rate* into location $j$ between $\tau - 5$ and $\tau$ is defined as the difference between inflow and outflow rates: $\widehat{x}_{j\tau} - \widehat{o}_{j\tau}$.

**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 and inflows defined above in the following way:

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

Group-specific outflow and inflow rates are then defined as

$$\widehat{o}_{jg\tau} = \frac{out_{jg\tau}}{\overline{y}_{jg\tau}}, \ \widehat{x}_{jg\tau} = \frac{in_{jg\tau}}{\overline{y}_{jg\tau}}.$$

Define the population share of cell $g \in G$ over the U.S. population:

$$\upsilon_{g\tau} = \frac{\sum_j \overline{y}_{jg\tau}}{\sum_j \sum_g \overline{y}_{jg\tau}}.$$

When aggregating the inflow and outflow measures across cells, I use the weight $\upsilon_{g\tau}$ to control for composition effects. So, the adjusted outflow and inflow rates for location $j$ are defined as:

$$\begin{aligned} \widehat{o}_{j\tau}^{adj} &= \sum_g \upsilon_{g\tau} \times \widehat{o}_{jg\tau}, \\ \widehat{x}_{j\tau}^{adj} &= \sum_g \upsilon_{g\tau} \times \widehat{x}_{jg\tau}. \end{aligned}$$

Similarly, it is possible to define net flows.

**Weekly Earnings**

Workers' weekly earnings were computed using data from the Census of Population. The same sample selection criteria listed above in the section on Sample Selection 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 preceeding 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.

**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 the Sample Selection section. 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.

# B The Benchmark Model

Under the assumption that the shock $\upsilon_1$ is sufficiently large, workers whose match is bad will always choose to migrate. Their value function is then given by:

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

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 $\upsilon$:

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

It is possible to simplify the expression for $u(s,\zeta,\upsilon)$ by using the functional form (13) for $\phi$ (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 $z$:

$$r(s,\zeta) = \left[\frac{z(1-\tau)^{1-\tau}(\tau)^{\tau}}{w(s,\zeta)^{\tau}}\right]^{\frac{1}{1-\tau}}. \tag{17}$$

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

$$u(s,\zeta,\upsilon) = w(s,\zeta)\left(1 + z^{-\frac{1}{\tau}}\left(\frac{A}{1-\tau}\right)^{\frac{1}{\tau}}\right) - \upsilon. \tag{18}$$

Land can be completely solved for by using the equilibrium conditions (10) and (11) together with the worker's first order condition for land (9). Replace the latter into equation (11) 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 efficiency units of 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 (17) to replace $r(s,\zeta)$ on the right-and side. This gives 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:

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

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 (18) and the unit wage is given by (19). With these preliminary steps, the model's equilibrium can then be characterized as follows. As far as the agents 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:

$$\int V(s',\zeta',\upsilon_2)\,Q(\zeta,d\zeta') \leq e-k. \tag{20}$$

  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 $\upsilon_2$. Thus, the components of $s'$ are:

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

Notice that $q(s,\zeta,\upsilon_2) \leq 1$ is implicitly defined by (20) in case of equality.

- **Region 2**. None of the $\upsilon_2$ workers chooses to leave and no new worker chooses to locate there. In this case:

$$\int V\left(s',\zeta',\upsilon_2\right)Q(\zeta,d\zeta') \;>\; e-k, \tag{21}$$

$$p\int V\left(s',\zeta',\upsilon_1\right)Q(\zeta,d\zeta') + (1-p)\int V\left(s',\zeta',\upsilon_2\right)Q(\zeta,d\zeta') \;<\; e_d. \tag{22}$$

The first inequality expresses the fact that it is better for a $\upsilon_2$ 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:

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

Notice that, using equation (**??**), the two inequalities (21) and (22) can be rewritten as:[27]

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

- **Region 3**. None of the $\upsilon_h$ workers chooses to leave and some new workers choose to locate there. In this case, equation (21) still holds, while equation (22) holds as an equality:

$$e < \int V\left(s',\zeta',\upsilon_2\right)Q(\zeta,d\zeta') = \frac{e_d - p\left(\int u(s',\zeta',\upsilon_1)Q(\zeta,d\zeta') + \beta\delta\left(e-k\right) + \beta\left(1-\delta\right)e_r\right)}{1-p}. \tag{23}$$

The endogenous components of $s'$ are:

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

where $x_d(s,\zeta)$ is implicitly defined by the equality condition in (23).

# 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:

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

[27] 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.

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

Solving the optimal choice of land and consumption yields the 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 $(a,\varepsilon)$.

Replacing the expressions for the wage and rent yields:

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

Notice that in this case there are no productivity shocks and wages are given by:

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

so 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 bechnmark 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)). Rappaport (2007) emphasizes changes in tastes for amenities for understanding the geographical reallocation of population in the U.S.

# 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 $\theta$, one for finding the equilibrium values $(\overline{x},e_d,e_r)$, of the model for given $\theta$, and one for finding $\theta$ in order to match the empirical moments of interest. Every change in $\theta$

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 $\theta$ 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 $\theta$ is updated in Step 4.

**Step 2 (Dynamic Programming).** Solve the dynamic programming problem described in Section (3). 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 $n$ while $z$ includes $z$ and $\varepsilon$) and because the procedure involves numerical integration of the value function with respect to the density of the innovation $u$. Last, it is necessary to take into account the possibility that the constraint that keeps gross inflows from becoming negative binds $(x \geq 0)$. The solution of the dynamic programming problem yields gross inflows $x(s,z)$ and the probability of outflow $q(s,z)$ for an agent with match $\upsilon_h$ as functions of the state vector $(s,z)$. These two functions allow one to recover all the other variables of interest, in particular the location's population $y(s,z)$ conditional on $(s,z)$.

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

$$f_1\left(\overline{x}, e_d, e_r\right) = \int x(s,z)\mu\left(ds, dz\right) - \eta\overline{x}, \tag{24}$$

$$f_2\left(\overline{x}, e_d, e_r\right) = \int y\mu\left(ds, dz\right) - 1, \tag{25}$$

$$f_3\left(\overline{x}, e_d, e_r\right) = p\int V\left(s,z,\upsilon_1\right) y\mu\left(ds, dz\right) + (1-p)\int V\left(s,z,\upsilon_2\right) y\mu\left(ds, dz\right) - e_r. \tag{26}$$

The value $(\overline{x}^*, e_d^*, e_r^*)$ such that $f\left(\overline{x}^*, e_d^*, e_r^*\right) = 0$ represents the equilibrium value of migration. The integral in equations (24)-(26) 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 (24)-(26) with the corresponding sample averages. For example, the integral in (25) is approximated by:

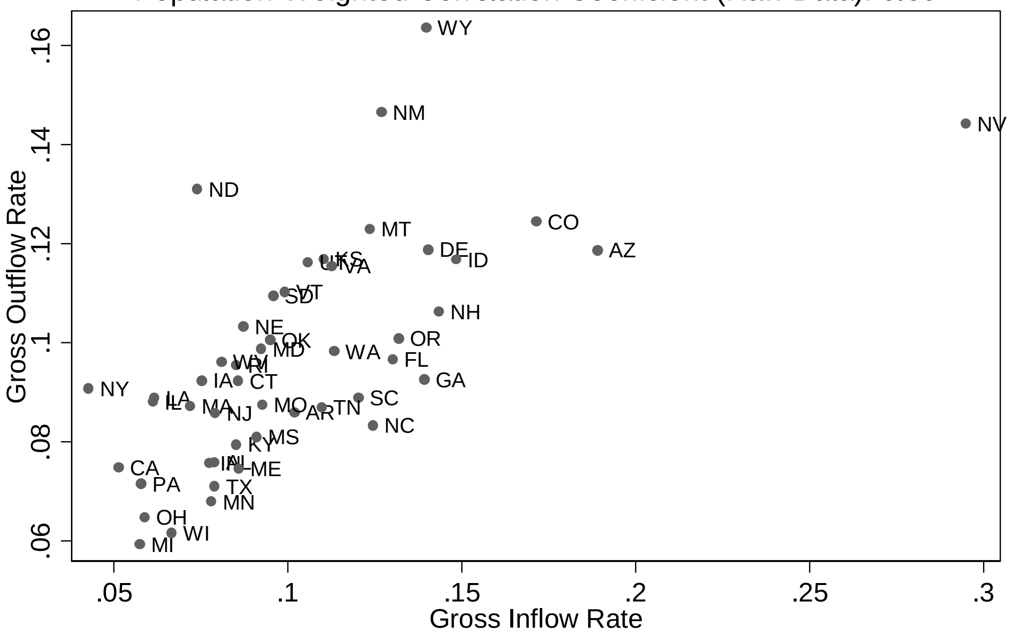
$$\frac{1}{T}\sum_{t=1}^{T} y_t.$$

The zero of (24)-(26) 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 $\left(\overline{x}^*, e_d^*, e_r^*\right)$, it is feasible to compute the equilibrium value of all the variables of interest. The vector $\theta$ is estimated by constructing the model counterpart of the six moments listed in the text (Section 4) and choosing $\theta$ so that the model-generated moments are exactly equal to their empirical counterparts. Since there are six parameters and six moments, this is an exactly identified model. The problem then becomes one of solving six non-linear equations in six 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 six equations in six 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 $\theta$ 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 backtracking along the Newton dimension. When an acceptable step is determined, $\theta$ 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 $\theta$ 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 $10^{-4}$.

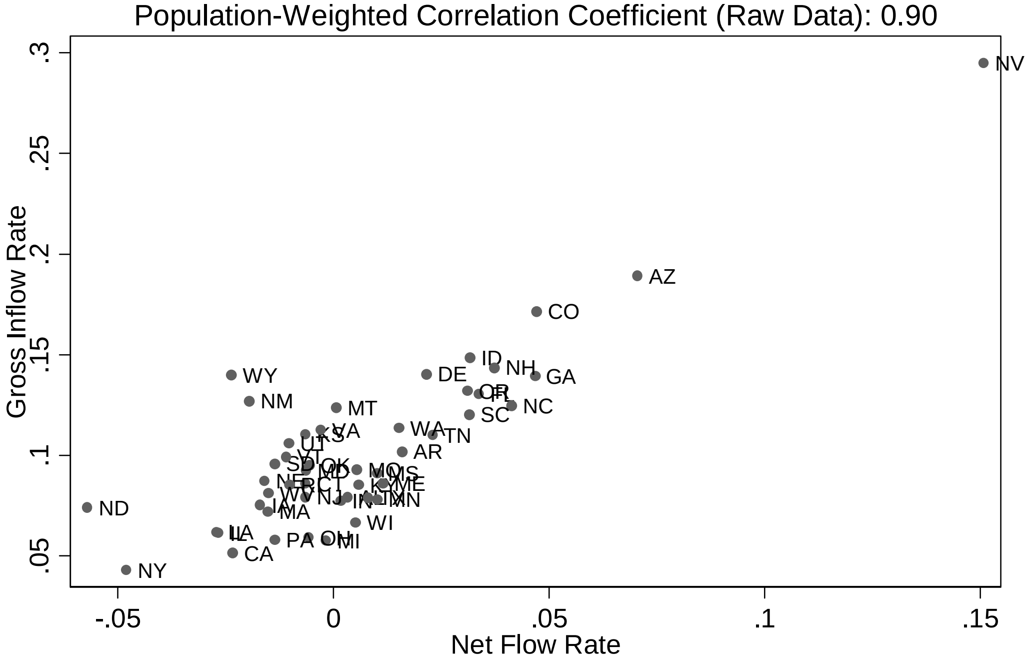
## Figure 1 - Gross Inflow and Outflow Rates (2000 Census)

Population-Weighted Correlation Coefficient (Raw Data): 0.69



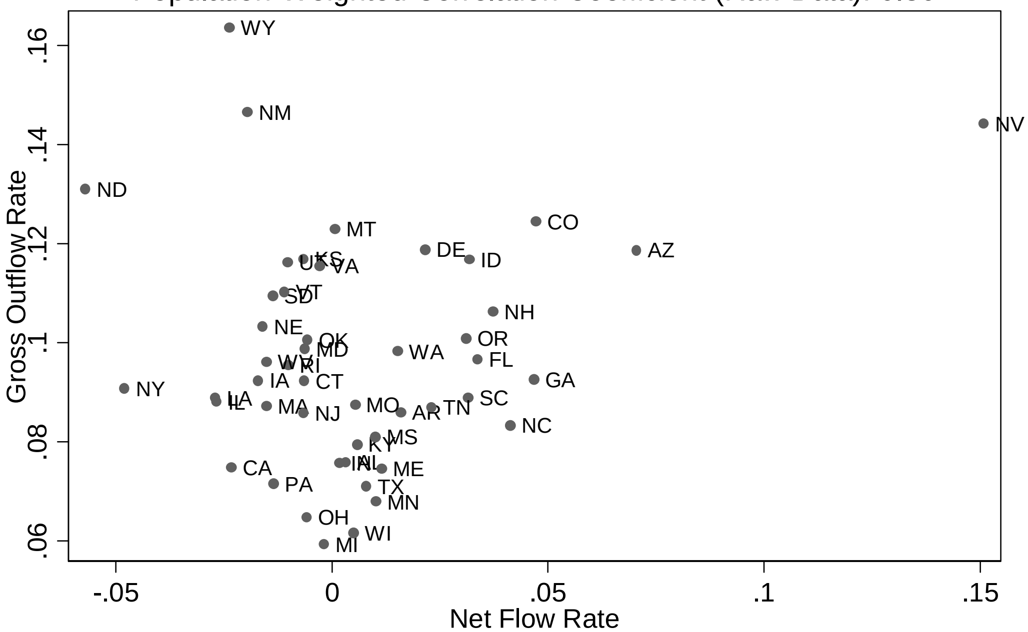

## Figure 2 - Gross Inflow and Net Flow Rates (2000 Census)

Population-Weighted Correlation Coefficient (Raw Data): 0.90



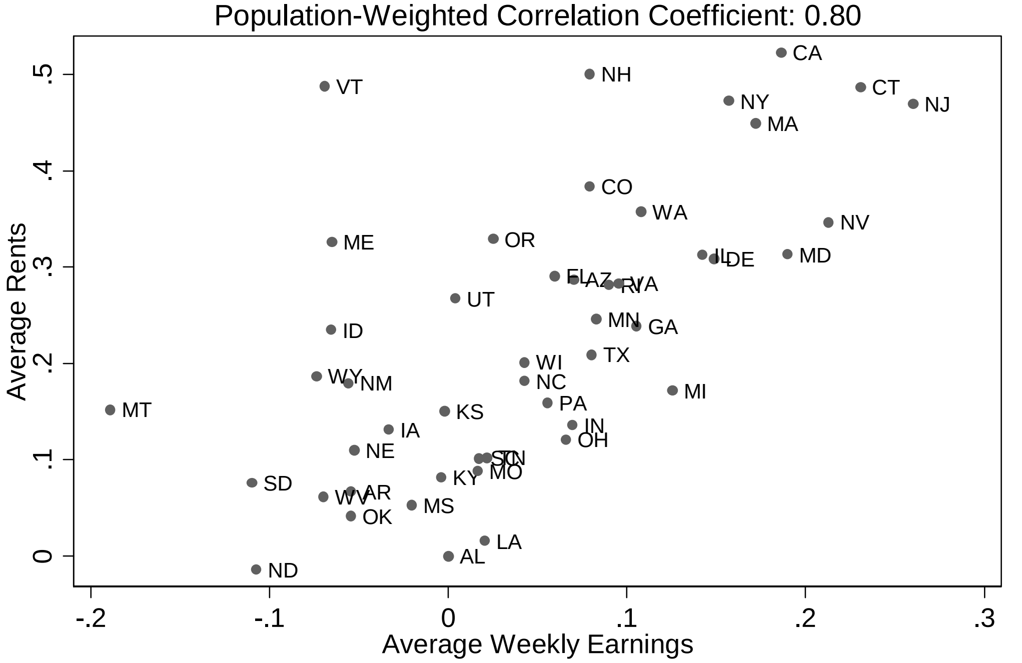
## Figure 3 - Gross Outflow and Net Flow Rates (2000 Census)

Population-Weighted Correlation Coefficient (Raw Data): 0.30



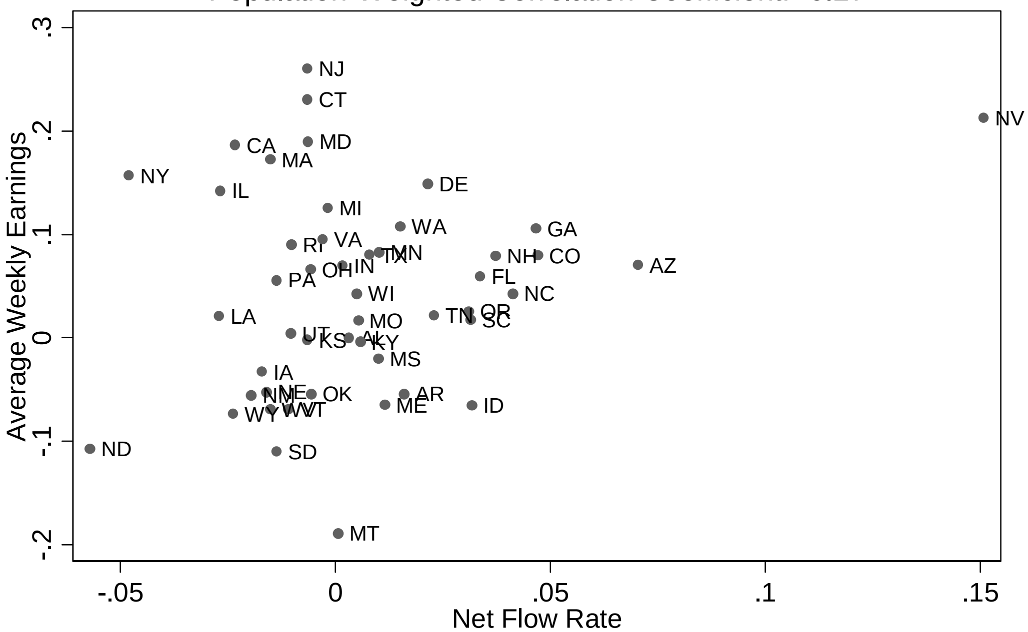

## Figure 4 - Average Weekly Earnings and Average Rents

Population-Weighted Correlation Coefficient: 0.80



## Figure 5 - Net Flow Rates and Average Weekly Earnings

Population-Weighted Correlation Coefficient: -0.27




## Figure 6 - Net Flow Rates and Average Rents

Population-Weighted Correlation Coefficient: -0.31

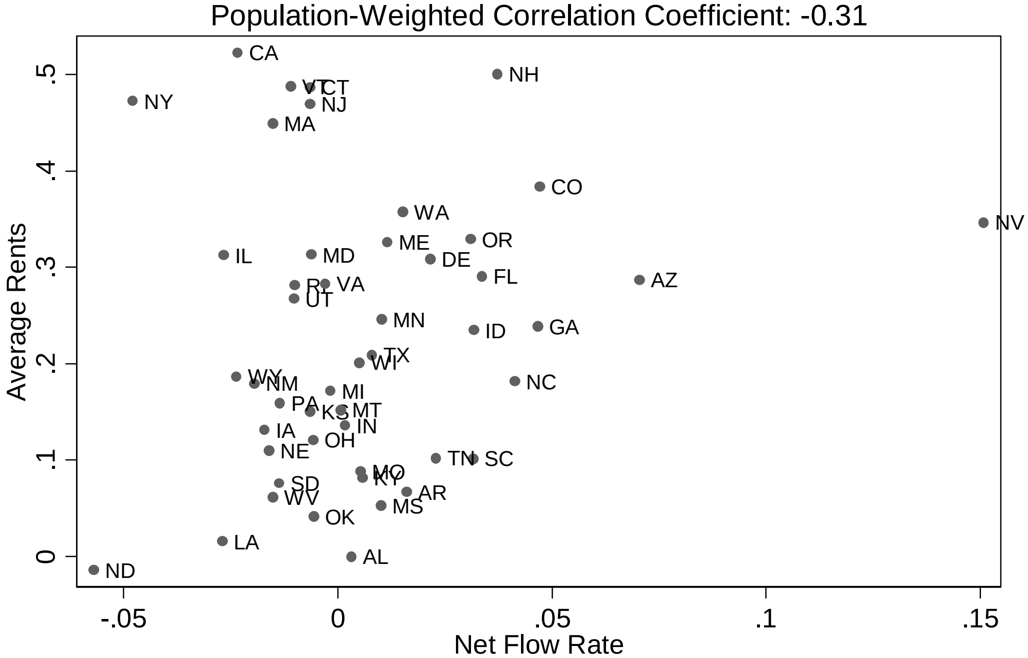


Figure 7: Scatter Plot of Outflow and Inflow Rates Against Net Flow Rates

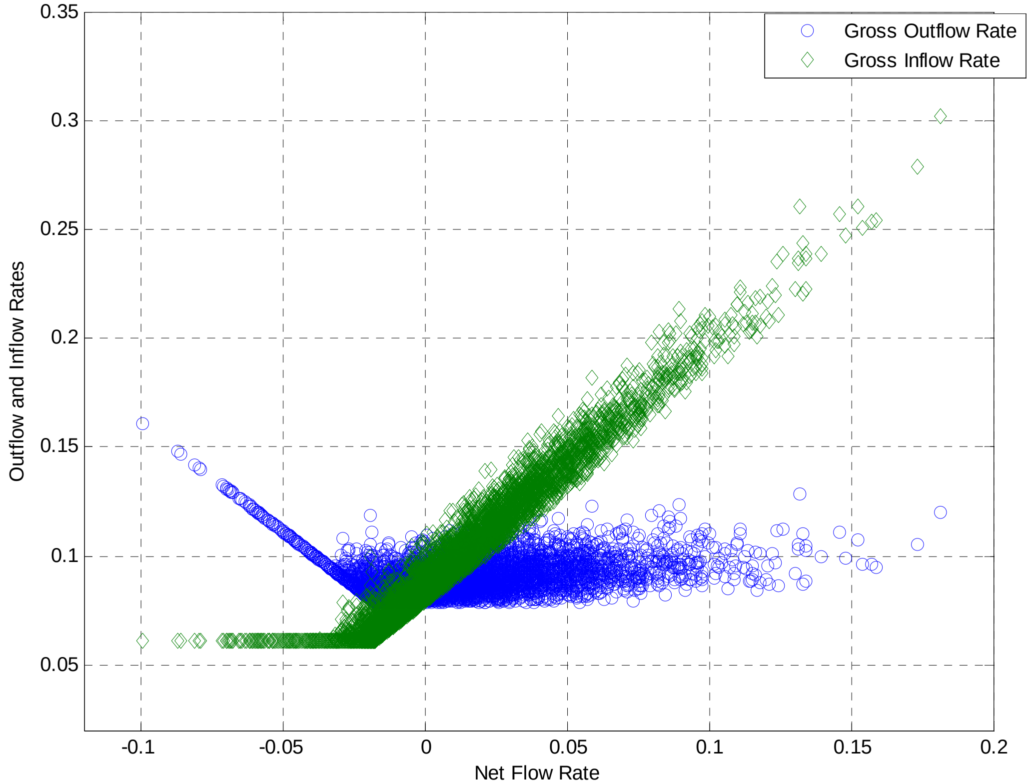


Figure 8: Scatter Plot of Location-Specific Wages Against Net Flow Rates

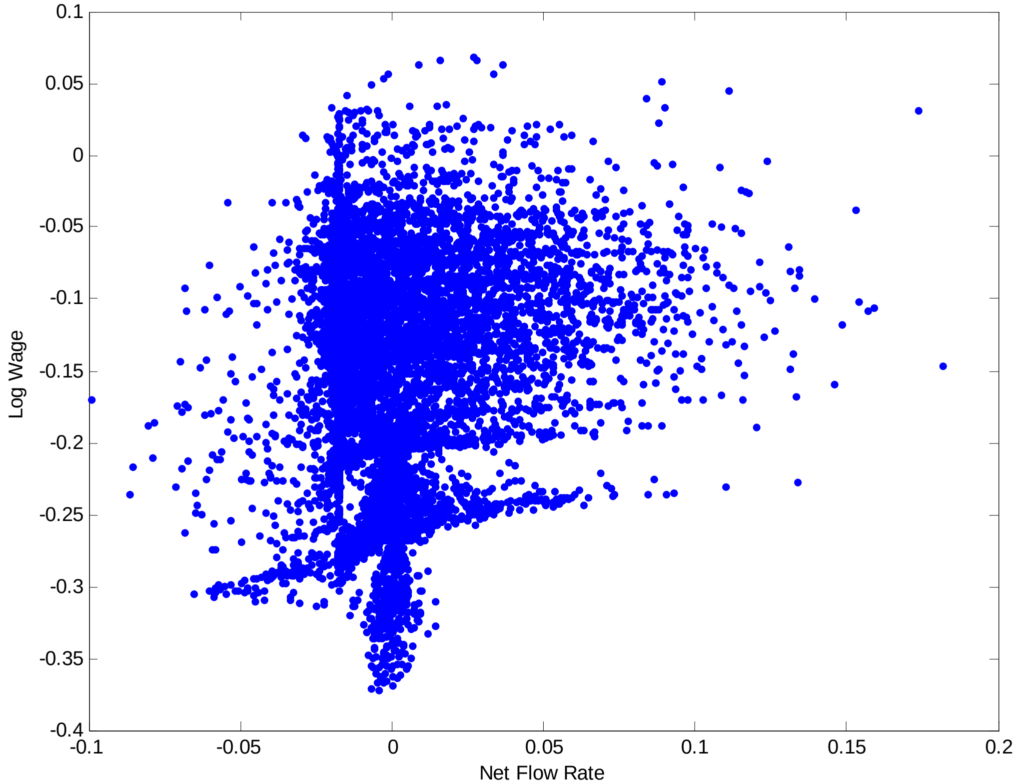


Figure 9: Scatter Plot of Location-Specific Rents Against Net Flow Rates

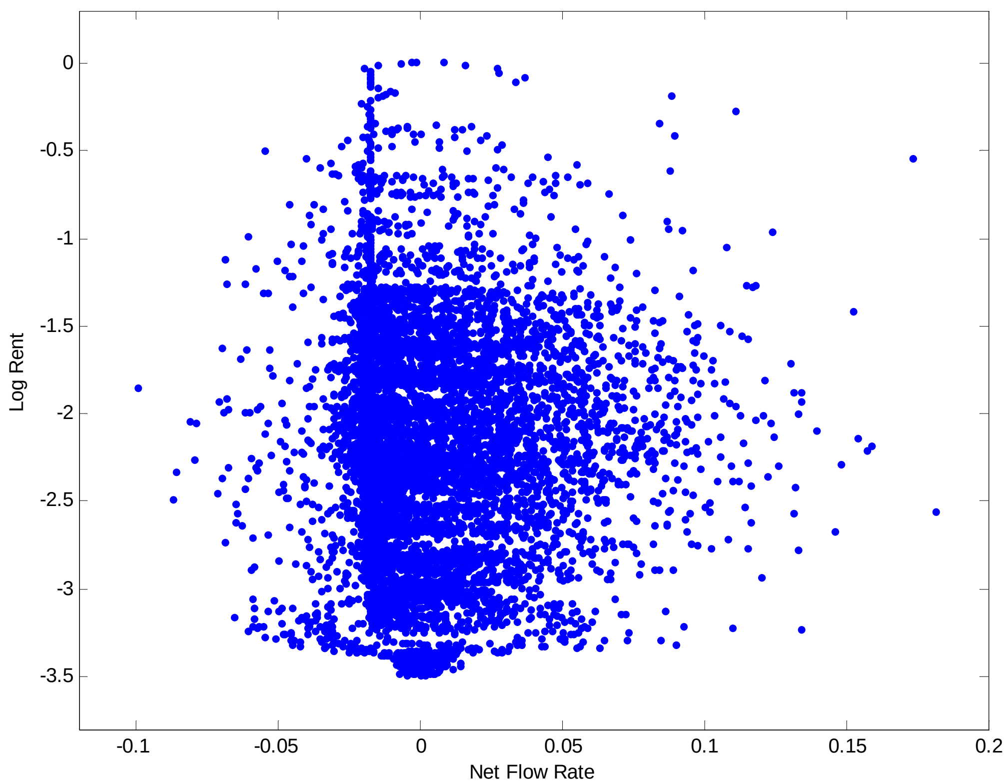


Figure 10: Scatter Plot of Location-Specific Rents Against Wages