Understanding Gross Workers Flows Across U.S. States

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

This Version: July 2006. First Version: February 2006.

Abstract

This paper documents and provides an explanation for the main stylized facts about net and gross workers flows across states in the U.S. While it is generally known that gross flows of population across locations are significantly larger in the U.S. than within most European countries, there is considerable heterogeneity in gross and net flows across states within the U.S. itself. The main purpose of the paper is to test whether a general equilibrium model based on Lucas and Prescott (1974)'s island economy, augmented to allow for gross workers flows, can account for the main stylized facts. The key stylized facts are as follows. 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 are uncorrelated. 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. To address these facts, I develop a general equilibrium model of net and gross workers' flows across locations. Net flows are driven by shocks to local labor demand, while gross flows are driven by idiosyncratic location-specific shocks to workers' productivity. In response to shocks to the growth rate of labor productivity in a location, the model generates artificial data that are generally consistent with the stylized facts listed above. Using the estimated parameters I find that the contribution of excess workers flows to aggregate welfare is about one percent of aggregate output in the benchmark economy.

*E-mail: coenp@andrew.cmu.edu. Thanks, without implication, to seminar participants at UT Austin, SUNY Albany, Cleveland Fed, LABORatorio Revelli and the 2006 Midwest Economics Association and 2006 Society for Economic Dynamics Meetings, and to Yoonsoo Lee, Chris Flinn, Holger Sieg, Fallaw Sowell, Marla Ripoll, Maria Ferreyra, Dennis Epple, Alexis Leon, Iourii Manovski, and especially Rui Castro, for useful discussions and suggestions. Richard Lowery provided able research assistance.

1 Introduction

Workers' flows across U.S. states are the main factor behind the large and persistent dispersion in states' relative employment growth over time (Blanchard and Katz (1992)). Yet, the process by which population is reallocated among geographic areas within a country is not well understood. 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 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. The second goal is to determine whether the main stylized facts can be explained using a general equilibrium model of workers flows across locations. The model is a version of Lucas and Prescott (1974)'s island economy, extended to allow for gross migration flows.

I start by documenting these facts using the decennial Census of the U.S. for the post-WWII period. 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. These flows are adjusted to take into account the different demographic and industrial composition of the workforce across states and differences in other state characteristics, such as size.

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 2.2 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 tend to be uncorrelated. These facts seem to suggest 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. 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, there is a significantly positive cross-sectional correlation between average state wages, adjusted by differences in living costs, and net flow rates.

In order to account for these facts, I consider a model of gross and net flows. The model economy is composed by a set of local labor markets ("islands"), that are hit by idiosyncratic labor demand shocks. Local wages would tend to rise in response to these shocks, but workers' mobility across islands tends to equalize the price of an efficiency unit of labor across islands. At a point in time, a location typically experiences both gross inflows and gross outflows. This is because a worker's idiosyncratic productivity differs across islands, giving rise to workers' gross flows. In general equilibrium, the value of migration is pinned down by a zero excess demand condition for aggregate net flows.

The model's parameters are estimated using a method of simulated moments. The estimated model is consistent with the main stylized facts mentioned above. The mechanics of the model can be better understood by considering an unanticipated positive shock to the growth rate of local labor productivity. On impact, the workers' net flow rate rises while the outflow rate remains virtually constant. This is because gross inflows of workers are expected to arbitrage away the temporarily higher unit price of labor in the location. In the following periods, outflows rise above their steady state value, as some of the location's newly arrived workers are ex-post unlucky and decide to move again. The persistent nature of innovations to local labor demand shocks implies that net flows remain above steady state for several periods. Due to the response of gross outflows, gross inflows exceed net inflows. Thus, gross inflows are more volatile than net inflows, which, in turn, are more volatile than gross outflows. Gross inflow and outflow rates are positively correlated as larger gross inflows of workers are followed by larger gross outflows.

I use the estimated model to assess the contribution to aggregate output of workers' excess flows (i.e., gross flows minus absolute net flows) across U.S. states. By counterfactually imposing that workers cannot migrate in order to improve their idiosyncratic match with their state of residence, I find that excess flows account for about one percent of aggregate output.

This paper is related to several literatures. The closest literature is the one initiated by Lucas and Prescott (1974) in their "island" model of the labor market and extended by Jovanovic and Moffitt (1990) to account for gross flows.[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 productivity shocks, giving rise to gross flows of workers.

The importance of gross flows of workers across sectors was first highlighted by Jovanovic and Moffitt (1990) who considered a simplified version of the Lucas-Prescott model allowing for sector-specific shocks to workers' productivity. An important insight of this 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 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 cannot reproduce the

[1] Topel (1986) considers a setting similar to Lucas and Prescott (1974).

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 inflows are zero.

The paper also builds on the contribution by Blanchard and Katz (1992), who developed 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.[2]

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 offers a discussion of some modelling issues. Section 5 evaluates the model in light of the stylized facts of Section 2. Section 7 computes the contribution to aggregate output of excess flows of workers across U.S. states. Section 8 concludes. The data appendix offers a more detailed description of the data and the construction of the flow variables.

2 Data and Stylized Facts

Data on Workers' Flows. The main data set I use is the U.S. Census of Population for several decades.[4] The Census data have the clear advantage of being a large and comprehensive dataset. 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.

Since 1940, the Census questionnaire has included a question regarding the location (state and metropolitan area) 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

[2] 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. The paper is also related to the traditional 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 workers flows.

[3] See also Dahl (2002) and Borjas, Bronars and Trejo (1992).

[4] This is available online at `www.ipums.org/usa`.

[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.

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. The appendix 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 the appendix.

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.[6] 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 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, according to the Census there are more than 200 metropolitan areas in the U.S. This figure makes the estimation of the model extremely lengthy, while it is feasible, yet long, to work with 48 locations. Last, for the purpose of policy analysis, many labor market policies (e.g. unemployment insurance) are set at the state level.

Composition Effects and Heterogeneity Across States. Figures 1-3 report scatter plots of outflow, inflow and net flow rates computed using the raw data from the 2000 Census. There is, of course, considerable heterogeneity among states in at least two dimensions. First, at the micro level, different states have a different composition of population, in terms of age, education, industry of employment, etc. When comparing measures of population flows across states, one has to make sure to control for possible composition effects. It can, in fact, be that certain states exhibit higher gross flows because of the sectoral or demographic composition of their employment structure. For example, if the gambling industry has a particularly high turnover of workers', then we might expect the state of Nevada, in which this industry is particularly large, to feature large inflows and outflows of workers. To address the micro heterogeneity, I divide the population into 490 demographic groups defined by age, education, and industry. Then, I compute gross outflow and inflow rates for each state and for each demographic group. Last, I compute 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 gross flows obtained using this procedure are very

[6] Specifically, the Census variable `migmet5` (metropolitan area of residence 5 years before) is not available in 1970, while the variable `migplac5` (state of residence 5 years before) is.

close to the unadjusted ones.[7] 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.

At the macro level, states have different sizes, different numbers of large metropolitan areas, etc. The concern here is that differences in gross flows might be driven by some of these factors, as opposed to the economic forces I would like to emphasize. To address this macro heterogeneity, I have run a cross-sectional regression of inflow and outflow rates (adjusted using the above procedure) on states' land area, number of metropolitan areas with population above 0.5 million, and year when the state formally joined the U.S. I have then defined the outflow and inflow rates to be the residuals of this regression, and the net flow rates as the difference between the two. This second adjustment has a more sizeable effect on the statistics of interest, but does not affect the basic properties of the data, either.

Inflows and Outflows in the Cross-Section. Tables 1 and 2 below provides descriptive statistics regarding inflow, outflow and net flow rates across U.S. states using data from the Census 2000, adjusted as described above.[8]

Table 1
Basic Statistics on Workers Flows (Census 2000)

	Mean	Median	Standard Deviation	Minimum	Maximum
Outflow Rate	8.86	8.67	1.54	5.46 (Wisconsin)	17.15 (Wyoming)
Inflow Rate	8.86	8.27	3.33	4.03 (North Dakota)	28.26 (Nevada)
Net Flow Rate	0.00	−0.10	2.59	−7.99 (North Dakota)	12.51 (Nevada)

Table 2
Cross-Sectional Correlations (Census 2000)

	Outflow Rate	Inflow Rate	Net Flow Rate
Outflow Rate	1	0.66***	0.25*
Inflow Rate		1	0.89***
Net Flow Rate			1

*** denotes significant at 1% level, * at 10% level

From these two Tables, some interesting facts emerge:

- Gross flows are large relative to net flows. Between 1995 and 2000 the average state gained or lost about 2.2 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 (Table 1).

[7] The cross-sectional correlation between adjusted and unadjusted rates in the 2000 Census is always above 0.97.
[8] The statistics in this and the following tables are computed weighting each state by its relative population.

- There is a relatively large dispersion across states in outflow, inflow and net flow rates, with outflow rates being relatively less dispersed than gross and net inflow rates (Table 1).

- States that experience a relative large gross inflow of population also tend to experience a relatively large gross outflow of population (Table 2 and Figure 1). For example, the state of Nevada ranked first in terms of gross inflows (about 28.26 percent) and second in terms of gross outflows (15.74 percent). Interestingly, the positive correlation between gross inflows and outflows is apparently a well-known, though not extensively documented, stylized fact in the literature on internal migration of population (see Greenwood, 1975).[9] The correlations in Table 2, particularly the one between gross inflows and gross outflows, are consistent with two polar situations. One in which the large flows are symptoms of a changing industry/demographic mix of the state's workforce, so that the outgoing workers are different from the incoming ones. The other in which the "same" type of worker moves in and out of the state. In order to distinguish between these two possibilities, I have used the 490 demographic groups (indexed by g and described above) and computed, for each state j and for the 2000 Census, the following measure of within demographic group workers' reallocation:[10]

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

  The denominator of this expression gives the difference between the sum of gross inflow and outflow from location j and the absolute net flow. Thus, it represents the excess of workers' mobility over and above what is needed to accommodate net workers' flows. To understand the numerator of equation (1), suppose that inflows and outflows of workers always occurred between demographic groups. This means that for each group g, we would either have $in_{jg} > 0$ and $out_{jg} = 0$ or $in_{jg} = 0$ and $out_{jg} > 0$. In this case the numerator of (1) would be zero, and so would the measure of within-group reallocation. At the other extreme, if inflows and outflows were always balanced within groups ($in_{jg} = out_{jg}$), then the index would be equal to one. The outflows-weighted average of this measure across states for the 2000 Census was 0.91, suggesting that most flows occur within the demographic/industry groups described above. 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 for each group separately. Then, the 490 correlation coefficients can be averaged using as weights the groups' population shares in the U.S. The following Table reports these adjusted

[9] See, for example, Miller (1967, page 1426, Table 3). She defines locations in terms of metropolitan areas, instead of states and shows, using 1960 Census data, that this correlation is robustly positive both in the aggregate and within demographic groups defined by sex, race and occupational category.

[10] 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.

correlation:

Table 3
Adjusted Cross-Sectional Correlations (Census 2000)

	Outflow Rate	Inflow Rate	Net Flow Rate
Outflow Rate	1	0.39	-0.23
Inflow Rate		1	0.79
Net Flow Rate			1

Comparing the correlations in Tables 2 and 3, one notices that the latter are smaller. The signs of the correlations between gross inflows and outflows are the same in both Tables, while the correlation between outflows and net flows turns negative in Table 3. These results are consistent with the view that some of the gross flows we observe have to do with changes in the composition of states' workforce. From our perspective, however, it is important that, even within narrowly defined demographic/industry groups, there is a positive correlation between gross inflows and outflows.[11]

- Gross and net inflows are highly and positively correlated in the cross-section, while the correlation between net flows and gross outflows is in absolute value smaller (Tables 2 and 3 and Figures 2 and 3). This observation, together with the previous two, seems to suggest 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, across states. In other words, states that tend to lose population to other states seem to do so by attracting fewer new workers as opposed to losing more local ones.

Cross-Sections Over Time and the Time-Series Dimension of Workers' Flows. It is natural to ask whether the statistics presented in the previous Tables are peculiar to the 2000 Census or not. It is also important to determine how much persistence there is in gross population flows for a given state. Both questions can be answered by considering other Census years.

Tables 4 and 5 confirm that the salient features of gross and net flows pointed out above in relation to the 2000 Census are also present in the 1970-1990 Censuses.[12,13]

[11] For consistency, one could also adjust the cross-sectional standard deviation of flow rates in Table 1 in order to capture exclusively the dispersion of flows within demographic groups, and not the cross-groups covariance terms. By doing so, since the latter covariances tend to be negative, one would obtain higher standard deviations for inflow, outflow, and net flow rate. Their ranking does not change, though. For simplicity, I do not carry out this further adjustment.

[12] 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 6 below. The 1940 Census does not present particular problems.

[13] For simplicity, given that the first type of adjustment mentioned above (for composition effects related to age, education and industry of employment) did not produce any sizeable effect on the statistics of Tables 1 and 2, in this Table the data for 1970-2000 are only subject to the second type of adjustment mentioned above. This explains the difference between the results in Table 1 and the one in this Table for the 2000 Census.

Table 4
Basic Statistics on Population Flows (Censuses 1970-2000)

	2000		1990		1980		1970	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Outflow Rate	8.59	1.50	9.04	2.01	8.97	1.77	6.89	1.72
Inflow Rate	8.59	3.22	9.04	3.90	8.97	3.86	6.89	2.73
Net Flow Rate	0.00	2.60	0.00	3.84	0.00	3.06	0.00	2.39

Table 5
Cross-Sectional Correlations (Censuses 1970-1990)

	Outflow Rate	Inflow Rate	Net Flow Rate
Outflow Rate			
1970	1	0.50***	-0.15
1980	1	0.63***	0.22
1990	1	0.29**	-0.23
2000	1	0.60***	0.17
Inflow Rate			
1970		1	0.78***
1980		1	0.89***
1990		1	0.86***
2000		1	0.89***

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

Do states with relatively high gross and net flows between 1995 and 2000, also tend to display relatively high flows between 1985 and 1990, and before? The answer to this question is affirmative for both gross and net flows. The following Table reports, for each type of flow, its autocorrelation coefficient across Census years, computed by pooling all state-year data points together.[14]

[14] 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 4.538. This number guarantees that the cross-sectional standard deviation of net flow rates in 1950 is the same as the average standard deviation in all the previous Census years. Notice that in Table 6 the difference between t and $t-1$ is equivalent to 10 years.

Table 6
Autocorrelations of Population Flows (Censuses 1950, 1970-2000)

	Autocorrelation Coefficient				
	$t, t-1$	$t, t-2$	$t, t-3$	$t, t-4$	$t, t-5$
Outflow Rate	0.66***	0.68***	0.29***	0.49***	0.59***
Inflow Rate	0.86***	0.71***	0.56***	0.73***	0.73***
Net Flow Rate	0.73***	0.61***	0.55***	0.41***	0.40***

*** significant at 1% level.

Notice that all these flow rates are very persistent over time.

Migration Motives. Not all moves of workers across states are motivated by economic reasons. The Census survey does not contain any question regarding a worker's reason for the change of residence. However, since 1999, the March CPS has asked this question. I have aggregated the different answers to this question into two categories, according to whether an interstate move is attributable or not to economic factors. The appendix provides more information about the criteria used for this assignment. About two thirds of all interstate moves that occurred between 1999 and 2003, according to the March CPS, were of the first kind.

Earnings. Using the 2000 Census I have constructed a measure of a worker's weekly earnings in 1999. These nominal figures were deflated using the ACCRA cost of living index, which measures the relative price of a given basket of goods and services in a number of U.S. cities.[15] Denote the real weekly earnings of worker i in 1999 by w_i. The Census data provides detailed information regarding a worker's demographic characteristics, occupation and industry. Let these observables be summarized by the vector x_i. In order to construct measures of average wages within states and of residual wage inequality, I have run the following regression for log earnings:

$$\ln w_i = \mu \times STATE_i + \omega \times MOVE_i + x_i\beta + \varepsilon_i, \tag{2}$$

where $STATE_i$ is a vector of dummy variables recording individual's i state of residence in 2000, while $MOVE_i$ is a dummy variable that equals one if worker i moved across state lines sometimes between 1995 and 2000 and zero otherwise.[16] The estimates of this regression reveal two interesting

[15] See the appendix for more detail on this index. Due to limitations in the availability of this index, the size of the sample used to carry out the computations of this section is about 50 percent of the original sample size used in previous sections. However, it still includes about 2.5 million observations.

[16] As Kennan and Walker (2005) point out, the residual ε_i can be interpreted as reflecting the influence of (at least) three orthogonal factors on a worker's earnings: a location match component (which varies across locations for the same worker, but is constant within locations), a worker's fixed effect (which is constant within and across locations for a given worker), and a transitory effect (which varies both within and across locations for a given worker). Location-specific effects on earnings are a key determinant of migration decisions, in addition to differences in average earnings across states. Of course, the cross-sectional nature of the Census data does not allow one to separately identify the three components of residual wages. Fortunately, Kennan and Walker (2005, page 11) provide such decomposition using NLSY data. According to their estimates, location effects explain about 16 percent of the overall inequality in residual earnings among young high school educated workers.

facts. First, controlling for observables, the standard deviation of weekly earnings across states is about 11 percent of average weekly earnings in the U.S.[17] Second, the estimate of ω in equation (2) suggests that the weekly earnings of workers that have moved across state lines in the 5 years preceding the 2000 Census are about 1.8 percent *lower* than the weekly earnings of observationally equivalent workers who did not move during that period and lived in the same location in 2000. This evidence is consistent with the view that migrating workers are on average less productive than non-migrating ones in their new location of choice. The model introduced in the next section is consistent with this evidence.

3 Model

The model presented in this section builds on the island-model of the labor market developed by Lucas and Prescott (1974). The force that drives the dynamics of the local labor market in the model is a persistent labor demand shock.[18] Relatively high labor demand shocks generate temporary increases in local wages that are then followed by net inflows of workers. Simultaneously, idiosyncratic wage 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.

The economy is populated by a continuum of measure one of locations (islands). A location is denoted by $j \in [0, 1]$. All locations are ex-ante identical. Ex-post, locations differ because they are hit by different labor demand shocks, denoted by z_{jt}, where t indexes time.[19] An agent i located in period t in a location j has earnings $\upsilon_{ij} w_{jt}$. The latter are the product of two components. First, υ_{ij} is an idiosyncratic component which, as in Kennan and Walker (2005), represents the efficiency units of labor with which the worker is endowed in the location. These efficiency units remain the same as long as the agent stays in the same location. Second, w_{jt} represents the unit price of an efficiency unit of labor in location j. It is common to all workers living in location j and it changes over time with the local labor market conditions. An agent can freely move across islands to improve his idiosyncratic match υ_{ij} and the unit price of labor w_{jt}.[20]

In detail, the sequence of events is as follows:

- An agent i is born in a location j at the end of period $t-1$.

[17] This figure is calculated as $std\left\{\exp\left(\widehat{\mu}_j\right)\right\}/E\left\{\exp\left(\widehat{\mu}_j\right)\right\}$. Notice that in computing this number I am imposing the same vector of observables in all states.

[18] The next section argues that the model would yield the same implications for workers flows if it were driven by shocks to local amenities.

[19] In this model I abstract from unemployment. The flow data from the Census describe interstate moves at five years intervals, so a model's period will represent five years. Given that most unemployment spells last only a few weeks, this would create problems in numerically implementing the model. Lkhagvasuren (2005) considers a model that explicitly allows for unemployment in order to explain the large observed cross-state differences in unemployment rates.

[20] Including an explicit moving cost in the model would be straightforward. The only reason why it is not included is that the Census data used in this paper would not allow me to identify this parameter.

- At the beginning of t, the agent draws the idiosyncratic location-match υ_{ij} from the two-point distribution (υ_l, υ_h), whose mean is normalized to one. The individual productivity shock υ_{ij} and the local unit price of labor w_{jt} determine his wage $w_{ijt} = \upsilon_{ij} w_{jt}$.
- The agent receives a utility flow w_{ijt}.
- With probability $1 - \delta$ the agent dies and is replaced by another agent that will start his life in the same location at the beginning of period $t + 1$. With probability $\delta < 1$ the agent survives into the next period.
- If the agent survives, he can then decide whether to stay in location j 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.
- At the beginning of period $t + 1$, if the agent had remained in the same location j in which he was living in t, he receives momentary utility $\upsilon_{ij} w_{jt+1}$. If the agent has moved to a new location j', he draws a new idiosyncratic location-match $\upsilon_{ij'}$ from the same distribution as new-born agents.

Production. Aggregate output in location j at time t, denoted by a_{jt}, is produced according to the following Cobb-Douglas production function:

$$a_{jt} = z_{jt} l_{jt}^{\tau}, \ \tau \in (0, 1), \tag{3}$$

where l_{jt} represents labor, measured in efficiency units and z_{jt} is a shock to the productivity of labor located in j at time t. This production function embeds the assumption that there is a fixed factor of production, e.g. land, that gives rise to decreasing returns to scale at the local level. This assumption is necessary to guarantee that each location has a finite population at each point in time. The production function (3) is consistent with two alternative (and extreme) assumptions about capital mobility. Specifically, physical capital can be thought of as either being permanently fixed in all locations or as being perfectly mobile across locations. In the latter case, physical capital would move to equalize its expected marginal product across locations. The specification (3) represents the reduced form taken by the production function after the optimal amount of physical capital has been replaced back into its original specification.

The stochastic process $\{z_{jt}\}$ is assumed to be stationary and takes the form

$$z_{jt} = z_{jt-1}^{\psi} \varepsilon_{jt}^{1-\psi}, \tag{4}$$

where $\psi \in (0, 1)$ and

$$\varepsilon_{jt} = \varepsilon_{jt-1}^{\rho} u_{jt}, \tag{5}$$

where $\rho \in (-1, 1)$ and u_{jt} is independent and identically distributed both over time and across locations. Let $q(u)$ denote the density of this shock and normalize its mean to one. The specification of the exogenous shocks in equations (4) and (5) is non-standard because it assumes that the innovations ε_{jt} in (4) are persistent, as opposed to being identically distributed over time as in standard specifications. Persistence in the process followed by ε_{jt} is necessary in order to generate persistent net flows in the Lucas-Prescott model. If ρ were equal to zero, net flows would be negatively autocorrelated over time, which is strongly at odds with the data.[21] The timing of the model is such that z_{jt} is realized at the beginning of period t, after period $t-1$'s migration decisions have been made. Thus, a migrating agent only knows the expected value of location-specific productivity shocks.

The efficiency units of labor l_{jt} located in j at time t are:

$$l_{jt} = y_{jt}\overline{v}_{jt},$$

where y_{jt} represents the measure of workers located in j at the beginning of t and $\overline{v}_{jt}$ denotes the efficiency units per worker in that location. Specifically:

$$\overline{v}_{jt} = v_l \lambda_{jt} + v_h \left(1 - \lambda_{jt}\right), \tag{6}$$

where λ_{jt} denotes the fraction of agents located in j at the beginning of t characterized by an idiosyncratic match value v_l.

Firms' optimization yields the market wage rate per efficiency unit of labor:

$$w_{jt} = \tau z_{jt} l_{jt}^{\tau-1}. \tag{7}$$

Workers' Flows. Let x_{jt} and o_{jt} denote respectively gross inflows and outflows occurring between t and $t+1$, and y_{jt+1} denotes the population located in j at the beginning of $t+1$. With this notation, we can write the law of motion of population in a location:

$$y_{jt+1} = y_{jt} + x_{jt} - o_{jt}. \tag{8}$$

By definition $x_{jt}, o_{jt} \geq 0$. In each period t, there are three categories of workers who *might* leave a location j: 1. $(1-\delta)\, y_{jt-1}$ workers born in j at the end of $t-1$, 2. immigrants who arrived in j at the end of $t-1$, 3. residents of j who were also living in j in $t-1$. Notice that the only reason why an agent with idiosyncratic shock v_h might leave a location j is to obtain a higher unit price of labor. Instead, an agent with idiosyncratic shock v_l might want to leave to improve his match and/or to get a higher price of labor. In what follows, I will show how, given the absence of mobility costs, an agent with shock v_l will always find it more convenient to leave a location than to stay. An agent with shock v_h instead will leave only if the price of labor in a location becomes low enough. In that case he will have to be indifferent between staying and leaving. Let q_{jt} denote the probability that an agent with match v_h leaves a location. Let n_{jt-1} denote the sum of gross inflows and new born workers:

$$n_{jt-1} = x_{jt-1} + (1-\delta)\, y_{jt-1}.$$

[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 this, I choose the simpler specification in which the shock process is characterized by persistent innovations.

Then, gross outflows from location j between t and $t+1$ are:

$$o_{jt} = \delta p n_{jt-1} + \delta \left(y_{jt} - p n_{jt-1} \right) q_{jt}, \tag{9}$$

with $q_{jt} \in [0,1]$.

Given that individuals with a low match υ_l always migrate, the measure λ_{jt} of agents located in j at the beginning of t and characterized by an idiosyncratic match value υ_l is simply:

$$\lambda_{jt} = \frac{p n_{jt-1}}{y_{jt}}. \tag{10}$$

Recursive Formulation of Individual Mobility Problem. The state vector for an agent located in a location j is given by

$$s = \left(y, n_{-1}, z, \varepsilon \right),$$

where I have dropped the subscript j for simplicity. Then, the value function of a worker characterized by idiosyncratic match υ with the location is given by:

$$V\left(s, \upsilon; e\right) = \upsilon w\left(s\right) + \beta\delta \max\left\{ E\left[V\left(s', \upsilon; e\right) | s \right], e \right\},$$

where the expectation on the right-hand side of the Bellman equation is taken with respect to the endogenous distribution of the state vector s. Using equations (6), (7) and (10), the unit price of labor is:

$$w\left(s\right) = \tau z \left[\upsilon_l p n_{-1} + \upsilon_h \left(y - p n_{-1} \right) \right]^{\tau - 1}.$$

Consider now, agents with a low idiosyncratic component υ_l. These agents will always want to migrate to improve their idiosyncratic match. To see this, notice that there is no cost of moving, so that the only reason to stay in a location for one of these agents is a high unit price of labor. Locations with the highest unit price of labor will attract migrating workers. The latter have an expected utility of moving into one such location equal to e. It follows that a worker with match υ_l living in such location must have an expected utility lower than e. This implies that he is better off migrating than staying in that location. Thus, the Bellman equation for this worker is:

$$V\left(s, \upsilon_l; e\right) = \upsilon_l w\left(s\right) + \beta\delta e.$$

As far as the agents for whom $\upsilon = \upsilon_h$ are concerned, following Lucas and Prescott (1974), we need to distinguish among three different cases:

- Case A. Some (or all) of these workers leave and some (or none) remain. In this case, the remaining workers must obtain at least as much as the departing ones. The latter obtain, in expectation, e. Thus, in this case:

  $$E\left[V\left(s', \upsilon_h; e\right) | s \right] \leq e, \tag{11}$$

  where the endogenous components of s' are:

  $$\begin{aligned} n &= \left(1 - \delta\right) y, \\ y' &= \begin{cases} y\left(1 - \delta q\left(s\right)\right) - \delta p n_{-1}\left(1 - q\left(s\right)\right) & \text{if (11) holds with equality,} \\ y\left(1 - \delta\right) & \text{if (11) does not hold with equality} \end{cases}. \end{aligned}$$

  Notice that $q\left(s\right) \leq 1$ is implicitly defined by (11) in case of equality.

- Case B1. None of the υ_h workers leave and no new worker arrives. In this case:

$$\begin{aligned} E\left[V\left(s', \upsilon_h; e\right)|s\right] &> e, &(12)\\ pE\left[V\left(s', \upsilon_l; e\right)|s\right] + (1-p)\,E\left[V\left(s', \upsilon_h; e\right)|s\right] &< e. &(13)\end{aligned}$$

  The first inequality expresses the fact that it is better for a υ_h 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:

$$\begin{aligned} n &= (1-\delta)\,y,\\ y' &= y - \delta p n_{-1}.\end{aligned}$$

  Notice that, since

$$E\left[V\left(s', \upsilon_l; e\right)|s\right] = \upsilon_l E\left[w\left(s'\right)|s\right] + \beta\delta e,$$

  the two inequalities (12) and (13) can be rewritten as

$$e < E\left[V\left(s', \upsilon_h; e\right)|s\right] < \frac{e\,(1-p\beta\delta) - p\upsilon_l E\left[w\left(s'\right)|s\right]}{1-p}.$$

- Case B2. None of the υ_h workers leave and some new workers arrive. In this case:

$$\begin{aligned} E\left[V\left(s', \upsilon_h; e\right)|s\right] &> e,\\ pE\left[V\left(s', \upsilon_l; e\right)|s\right] + (1-p)\,E\left[V\left(s', \upsilon_h; e\right)|s\right] &= e. &(14)\end{aligned}$$

  The endogenous components of s' are:

$$\begin{aligned} n &= (1-\delta)\,y + x\,(s),\\ y' &= y + x\,(s) - \delta p n_{-1},\end{aligned}$$

  where $x\,(s)$ is implicitly defined by equation (14).

Thus, a value function $V\,(s, \upsilon_h; e)$ is a solution to the agent's problem if it satisfies:

$$V\,(s, \upsilon_h; e) = \upsilon_h w\,(s) + \beta\delta \max\left\{E\left[V\left(s', \upsilon_h; e\right)|s\right], e\right\},$$

where s' is such that

$$\begin{aligned} n &= (1-\delta)\,y + \max\left\{\widetilde{x}\,(s)\,, 0\right\},\\ y' &= y + \max\left\{\widetilde{x}\,(s)\,, 0\right\} - \delta p n_{-1},\end{aligned}$$

and $\widetilde{x}\,(s)$ is the solution to the following equation:

$$E\left[V\left(s', \upsilon_h; e\right)|s\right] = \frac{e\,(1-p\beta\delta) - p\upsilon_l E\left[w\left(s'\right)|s\right]}{1-p} \qquad (15)$$

when s' is such that

$$\begin{aligned} n &= (1-\delta)\, y + \widetilde{x}\,(s)\,, \\ y' &= y + \widetilde{x}\,(s) - \delta p n_{-1}. \end{aligned}$$

In words, if $\widetilde{x}\,(s)$, implicitly defined by condition (15), is negative, it must be that there is no incentive for migrating workers to choose this location because their expected utility upon receiving the shock υ_h (left-hand side of 15) would fall short of what is required to induce a positive gross flow (right-hand side of 15). In this case nobody enters the location.

After having found the value function that solves the problem of a worker with shock υ_h, one can use equation (15) to solve for gross inflows:

$$x\,(s) = \max\left\{\widetilde{x}\,(s)\,, 0\right\}.$$

Similarly, one can solve for the probability $q(s)$ that a worker with shock υ_h leaves a location. In particular, $q(s) = 1$ if

$$E\left[V\left(s', \upsilon_h; e\right)|s\right] < e,$$

when

$$\begin{aligned} n &= y\,(1-\delta)\,, \\ y' &= y\,(1-\delta)\,. \end{aligned}$$

Instead, if (11) holds with equality:

$$q\,(s) = \max\left\{\widetilde{q}\,(s)\,, 0\right\},$$

where $\widetilde{q}\,(s)$ solves:

$$E\left[V\left(s', \upsilon_h; e\right)|s\right] = e,$$

for

$$\begin{aligned} n &= (1-\delta)\, y, \\ y' &= y\,(1-\delta\widetilde{q}\,(s)) - \delta p n_{-1}\,(1-\widetilde{q}\,(s))\,. \end{aligned}$$

Stationary Equilibrium. The solution to the mobility problem in a location determines the law of motion of the endogenous variables of the state vector s. These laws of motion define the probability, denoted by $\Psi\,(y', n|s)$, that next period's population in the location is less than y' and gross inflows are less than n, given a state s this period. A stationary distribution of s is then a cdf $\Phi\,(y, n_{-1}, z, \varepsilon)$ that satisfies the following recursive equation:

$$\Phi\left(y', n, z', \varepsilon'\right) = \int\int\int\int I\left(z^{\psi}\left(\varepsilon'\right)^{1-\psi}, z'\right) Q\left(\frac{\varepsilon'}{\varepsilon^{\rho}}\right) \Psi\left(y', n|y, n_{-1}, z, \varepsilon\right) \phi\left(y, n_{-1}, z, \varepsilon\right) dy dn_{-1} dz d\varepsilon,$$

where $I(.,.)$ is the following indicator function:

$$I\left(z^{\psi}\left(\varepsilon'\right)^{1-\psi}\right) = \begin{cases} 1 \text{ if } z^{\psi}\left(\varepsilon'\right)^{1-\psi} \leq z' \\ 0 \text{ else} \end{cases},$$

and Q is the cdf of the shock u.

The stationary distribution Φ depends on the endogenous variable e. The expected utility of migration must be such that the sum of population across locations equals the aggregate population:

$$\int y\phi_y(y)\,dy = 1, \tag{16}$$

where ϕ_y denotes the marginal density of y.

4 Discussion

Before proceeding it is worth discussing some of the modelling choices that I have made.

Demand vs. Supply Shocks. The first modelling choice is that the driving force of the model is represented by local labor *demand* shocks, as opposed to local labor *supply* shocks. In doing so, I am not considering the possibility that the high net inflows of population experienced by states such as Nevada and Arizona in recent decades, might be driven by local amenities (e.g. warm winters). It turns out that this choice, while being empirically supported, is not restrictive. It is in fact possible to add amenities to the model of section (3) and show that it gives rise to identical implications for gross and net flows as the model driven exclusively by demand shocks. To do this, let a location j be characterized at time $t+1$ by a vector of amenities $\mathbf{k}_{jt+1}$ which can vary stochastically over time.[22] Assume that amenities affect agents' utility in a multiplicative fashion. An agent with match shock υ that lives in a location where the wage per efficiency unit of labor is w_{jt} has instantaneous utility $\upsilon w_{jt}h(\mathbf{k}_{jt})$. It follows that in this setting $h(\mathbf{k}_{jt})$ can play the same role as the labor demand shock z_{jt}.

The upshot of this discussion is that the model with amenities produces qualitatively identical implications for workers' flows as the model driven by labor demand shocks. In order to solve this identification problem, it is necessary to consider the relationship between real earnings and net flows. If the driving force of workers' flows were amenity shocks, one would expect a negative cross-sectional correlation between real earnings and net flows of workers.

The available evidence seems, *prima facie*, to suggest against this hypothesis. Table 7 reports the cross-sectional correlation coefficients among the following state-level variables: gross inflows, gross outflows, net flows, average (log) real weekly earnings (these four variables are from the 2000 Census), annual heating and cooling-degree days.[23]

[22] For example, the introduction of air conditioning has significantly improved living conditions in the South-West of the U.S. during summer months. This can be interpreted as a change in $\mathbf{k}$.

[23] I have used the dummy coefficient μ_j as a measure of average (log) weekly earnings in state j in 1999. The annual number of cooling and heating degree-days are from the U.S. Historical Climatography Series 5-2 and 5-1. They are averages over the period 1931-2000. Given that there might be considerable within-state variation in temperature, the series are constructed using a population-weighted aggregate temperature for each state, with weights given by the Census 2000 population. The number of annual cooling and heating degree-days for a state j in year t are defined

Table 7
Cross-Sectional Correlations (Census 2000)

	Outflow Rate	Inflow Rate	Net Flow Rate	Average Log Wages	Cooling Degree Days	Heating Degree Days
Outflow Rate	1	0.66^{***}	0.25^{*}	-0.17	0.25^{*}	-0.31^{**}
Inflow Rate		1	0.89^{***}	0.25^{*}	0.46^{***}	-0.47^{***}
Net Flow Rate			1	0.43^{***}	0.44^{***}	-0.42^{***}
Average Log Wages				1	0.37^{***}	-0.25^{*}
Cooling Degree Days					1	-0.81^{***}
Heating Degree Days						1

*** significant at 1% level, ** significant at 5% level, * significant at 10% level.

The Table clearly show how states with larger positive net flows of workers tend to be characterized by relatively higher real earnings (Figure 4).[24] These are also the states with better amenities, as measured by the number of cooling and heating degree-days (Figure 5). These correlations are, *prima facie*, inconsistent with theories that postulate the existence of compensating differentials in earnings across locations (see e.g., Roback (1982)).

Of course, the question of why certain states experience large positive and persistent labor demand shocks remains open. No attempt is made to answer this difficult question in this paper. The approach taken here will be to first use these net flows to back out the underlying local labor demand shocks. These shocks will then be fed into the model and the latter will be evaluated on the basis of its cross-sectional implications for the gross flows of workers.

The Land Market. It is simple to modify the economy of section (3) to include a housing market. Assume, for example, that agents care about the consumption good c and housing services h, according to the instantaneous utility function:

$$u(c,h) = c^{\alpha} h^{1-\alpha}.$$

The budget constraint of an agent located in j with match υ_{ij} becomes:

$$c + p_{jt} h = w_{jt} \upsilon_{ij},$$

where p_{jt} denotes the price of a unit of housing services. The latter are equal to the amount of land the agent decides to rent. Land is also used in production. Firms demand land to the point where

as:

$$\begin{aligned}\text{cooling degree days} &= \sum_{d=1}^{365} \max\{x_{jtd} - 65, 0\},\\ \text{heating degree days} &= \sum_{d=1}^{365} \max\{65 - x_{jtd}, 0\},\end{aligned}$$

where x_{jtd} is the average daily temperature in day d of year t in state j.

[24] A similar conclusion has also been reached by Topel (1986) and Blanchard and Katz (1992), who find it more support for local shocks to labor demand as the driving force behind net flows of workers across states.

its marginal product equals p_{jt}. In equilibrium, p_{jt} must be such that the demand for land for residential and business purposes equals its fixed supply. This version of the model can be solved along the same lines of the benchmark. Ignoring binding constraints on gross inflows of workers, for simplicity, the model with land would be characterized by the following key property. Agents would move across locations to equalize the expected value of unit labor prices w_{jt} deflated by a measure of the price level. For example, solving the agent's static optimization problem above and replacing his optimal choices back in the utility function yields:

$$\frac{w_{jt}v_{ij}}{p_{jt}^{1-\alpha}}.$$

Net flows across locations would then equalize the expected value of $w_{jt+1}/p_{jt+1}^{1-\alpha}$ across locations. In this version of the model net and gross flows would have the same properties as in the benchmark.[25]

Persistent Differences Across States. A crucial aspect of the modelling strategy concerns the nature of heterogeneity across states. The model above describes a *stationary* economy in which all locations are ex-ante *identical.* Differences across locations are persistent but not permanent. In particular, a location's population share and earnings per capita tend to return over time to their long-run value, which is common across locations. This modelling choice, although convenient for many reasons, deserves some further comment.

Consider first the evidence concerning per-capita income or wages. Barro and Sala-i-Martin (1992) and (1991), using state-level data dating back to 1840, have shown how states with lower initial incomes per capita have subsequently grown faster towards a common income level than initially richer states.[26] Blanchard and Katz (1992) also find strong evidence of convergence of manufacturing wages across U.S. states in the post-WWII period. Differently from Barro and Sala-Martin, though, in their empirical and theoretical analysis, Blanchard and Katz specify states' relative wages as stationary processes around state-specific means.[27]

Blanchard and Katz (1992, page 5) also document that, for the period 1950-1990, "U.S. states have experienced large and sustained differences in employment growth rates." They capture this observation in a formal way by modelling the growth rates of states' employment shares as stationary processes with state-specific means. Of course, since a state's employment share is bounded from above by one this formalization cannot be taken literally, but just as a convenient way of capturing the persistent differences in employment growth across states. This specification is not problematic for Blanchard and Katz because, in their reduced form model, they never impose the condition that the employment shares must be smaller than one. In the model above, instead, employment shares, rather than their first differences, are assumed to be stationary. This is a necessary assumption for a well-defined equilibrium. Of course, it is always possible to set the parameters controlling the degree of mean reversion to a small enough value, so that in practice it might be impossible to

[25] In principle, the model's implications for land prices and nominal wages could be separately tested, a step that might be taken in future research.

[26] To be more precise, Barro and Sala-i-Martin's evidence points to what is usually called in the growth literature "β-convergence". The latter denotes the tendency for states with lower incomes per capita to grow faster than states in which income per capita is relatively large.

[27] It is worth noticing that both Barro and Sala-i-Martin and Blanchard and Katz cannot control for differences in price levels across states.

distinguish one specification from the other. In summary, both with respect to relative wages and employment shares, I interpret observed differences across states as stemming from slow transitions, rather than from time-invariant features.

Firms' Mobility vs. Workers Mobility. The model of the previous section abstracts from capital. This is without loss of generality. It is easy to introduce perfect capital in the production function. The latter would then imply the equalization of the marginal product of capital across locations. This marginal product condition can then be used to solve for the amount of capital in each location as a function of its efficiency units of labor. The location's capital stock can be replaced in the original production function, leading to a new production function that depends only on labor.[28]

The Labor Market. In the model of the previous section, two types of agents are active in the labor market at a point in time. The first one is young individuals that are "just born", and draw their first wage shock. The second is movers from other locations who draw new wage shocks. As in Kennan and Walker (2005), an important assumption here is that there is no search at the local level, i.e., workers sample from the local wage offer distribution only once. This is a way to capture the fact that there is a local component to wages. For example, a company's manager might realize that he cannot advance in his career in the current location, but can move within the company to a different location, be promoted and given a salary raise. In the model the worker has to move to find out what his wage will be in the new location. While this is unlikely to be true in most cases, the worker is moving to a new job and there is going to be uncertainty regarding the job's features (colleagues, working hours, possibility of further advancements, etc.). This uncertainty generates uncertainty in wages, as the latter might reveal ex-post to be too low, given the job's characteristics.

5 Empirical Implementation

The model's parameters are estimated using the method of simulated moments (see Lee and Ingram, 1991 and Duffie and Singleton, 1993). When bringing the model to the data, one has to keep in mind that the model assumes the existence of a continuum of locations, and therefore a constant value of migration e. The empirical and artificial moments were, instead, constructed using data from 48 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

[28]Lee (2004) provides comprehensive evidence regarding plant relocation in the manufacturing sector for the period 1972-1992. He finds (page 17) that, in a five-year period, "plant relocations account for about 7 percent of variations in net employment growth across states. The remaining 93 percent is accounted for by within-state changes such as employment growth within continuing plants, de novo plant openings, permanent closings without relocation, and intrastate plant relocation." This evidence suggests that movements of existing plants play a relatively minor role in explaining the differential growth of states' employment. It does not rule out, though, that the opening of new plants could play a much more significant role. Blanchard and Katz (1992) provide evidence that suggests that neither plant relocation nor new plant openings play an important role in states' adjustment to local labor demand shocks. They estimate local labor demand equations and find that local employment reacts strongly to local labor demand shocks, while job creation responds weakly to movements in local wages.

state vector s across all other locations when solving his dynamic programming problem. In other words, e would become a function of the distribution of state vectors s across states. In practice, this discrepancy between the theory and the empirical implementation of the model is unlikely to be of any importance for two reasons. First, in the data the cross-sectional properties of gross and net flows tend to be very similar across Census decades, as predicted by the model with a continuum of locations. Second, and more important, the moments computed using the data generated by the model are robust to an increase in the number of locations above 48.

A period in the model is taken to represent 5 years. The discounting parameters β and δ and the production function parameter τ are set a-priori. The discount factor β is set equal to 0.82, implying a yearly interest rate $1/\beta$ 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 δ equal to 0.83, so that the average lifetime for an individual in the model is approximately 6 periods. Finally, the parameter τ is set to 2/3, which corresponds to the income share of labor in Gross Domestic Product.

The remaining parameters are $(\rho, \psi, \upsilon_l, p)$ and the parameter of the density $q(u)$. The latter is taken to be lognormal with mean one and variance $\exp\left\{\sigma_u^2\right\} - 1$. The parameter vector $\theta = (\rho, \sigma_u, \psi, \upsilon_l, p)$ is estimated by matching five moments constructed from the Census data.[29] The following table summarizes the model's parameters and their estimated values:

Table 8
Model's Parameters

Parameter	How it is set	Estimate
β discount factor	a-priori	0.820
δ probability of death	a-priori	0.830
τ labor share in output	a-priori	0.670
ψ mean-reversion parameter for z_{jt}	estimated	0.992
ρ first order autocorrelation of ε_{jt}	estimated	0.950
σ_u volatility of labor demand shock	estimated	0.300
υ_l value of low idiosyncratic shock	estimated	0.968
p probability of low idiosyncratic shock	estimated	0.364

The identification of the parameters is discussed below. First, ρ, σ_u and ψ are estimated by matching some of the cross-sectional and time-series moments of states' net inflow rates. Using four decades of Census data (1970-2000), I construct a panel of 4 observations on net inflow rates for each of the 48 contiguous U.S. states. I then consider the following three moments: the cross-sectional standard deviation of these net inflow rates in the 2000 Census (0.0259) and the first and second order autocorrelation coefficients of net inflow rates across Census years (0.73 and 0.61,

[29] The parameter υ_h is determined from p and υ_l from the normalization setting the unconditional mean of υ to zero. The appendix provides a detailed description of the numerical algorithm followed to estimate the parameters and solve the model.

respectively).[30] When constructing the model counterpart of these moments, it is important to consider the fact that the Census data are only available every ten years, while a model's period is equivalent to five years. Thus, in the estimation algorithm, the net inflow rates predicted by the model must be sampled every two model-periods.

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 8.86 percent.

Last, the parameter υ_l determines the relative earnings of migrants and incumbent workers in a location. Given that workers with low matches always leave a location, the incumbents are all characterized by a high match υ_h. Instead, a fraction p of incoming workers is characterized by a match υ_l and a fraction $1-p$ by υ_h. The parameter υ_l is identified by the estimated coefficient $\widehat{\omega}$ in the log wage regression (2). Given the simple structure of the model this parameter can be compute analytically. In fact, from section (2), the wage of migrants was about 1.8 percent percent lower than the wage of stayers in the location of destination. Thus:

$$\frac{p\upsilon_l + (1-p)\,\upsilon_h}{\upsilon_h} = 0.982. \tag{17}$$

Recall that the average efficiency units of incoming workers (the numerator of equation 17) have been normalized to one. Thus, we obtain υ_h and υ_l:

$$\begin{aligned} \upsilon_h &= 0.982^{-1}, \\ \upsilon_l &= \frac{1-(1-p)\,0.982^{-1}}{p}. \end{aligned}$$

6 Results

The following Tables represents the main cross-sectional and time-series statistics produced by the model. For clarity, I mark in bold the moments that were *not* targeted in the estimation of the model:

Table 9
Basic Statistics

	Mean		St. Deviation		Corr. Outflow		Corr. Inflow		Corr. Av. Wage	
	Data	Model	Data	Model	Data	Model	Data	Model	Data	Model
Outflow Rate	8.86	8.86	**1.54**	**0.84**	1	1	**0.39**	**0.88**	**−0.17**	**0.06**
Inflow Rate	8.86	8.86	**3.33**	**3.28**	**0.39**	**0.88**	1	1	**0.25**	**0.11**
Net Flow Rate	0.00	0.00	2.59	2.59	**−0.23**	**0.81**	**0.79**	**0.99**	**0.43**	**0.12**

[30] The first order autocorrelation coefficient has been computed by calculating the correlation between the vector of net flow rates in 2000, 1990 and 1980 and the lagged vector containing the net flow rates in 1990, 1980, and 1970. The second order autocorrelation coefficient has been computed in an analogous way.

Table 10
Autocorrelations of Population Flows

	Autocorrelation Coefficients					
	$t, t-1$		$t, t-2$		$t, t-3$	
	Data	Model	Data	Model	Data	Model
Outflow Rate	**0.66**	**0.68**	**0.68**	**0.55**	**0.29**	**0.46**
Inflow Rate	**0.86**	**0.82**	**0.71**	**0.66**	**0.56**	**0.57**
Net Flow Rate	0.73	0.73	0.61	0.61	**0.55**	**0.51**

The simulated model is generally consistent with the main stylized facts reviewed in section (2). In particular, in the cross section, gross inflows are predicted to be more dispersed than net inflows while the latter are predicted to be more dispersed than gross outflows. In addition, gross inflows and outflows and gross and net inflows are positively correlated in the cross section. However, the model incorrectly predicts that gross outflows should be positively correlated with net flows. In the time-series dimension all flows are very persistent. Last, states characterized by positive net flows also tend to have relatively high average wages, even if the model understates the magnitude of this correlation. In order to gain a better understanding of the working of the model, it is useful to consider the impulse response functions of workers flows to a labor productivity shock.

The Mechanics of Workers' Flows. Figure 6 plots the period-by-period response of outflow, inflow and net flow rates (represented as deviations from their steady state values) to a one-time unanticipated shock $u_{j1} > 1$. In detail:

- In period 1 gross outflows stay constant at their steady state level, while at the end of period 1, there is a net inflow of workers who are attracted by the expected positive conditions of the local labor market for period 2.[31] Thus, gross inflows increase by exactly the same amount as net inflows.
- In period 2, the shock u_{j2} is back to 1. Net flow rates, however, are driven by the expected growth in z_{jt}. Given that the growth rate shock ε_{jt+1} is positively autocorrelated, net flows in period 2 are also above their steady state value. Gross outflows tend to rise above steady state due to the relatively large gross inflow of workers in the previous period. Some of these workers, in fact, are ex-post unlucky (i.e. they draw an idiosyncratic shock v_l) and decide to migrate again. Notice that the incoming workers leave the location at the same rate as individuals that are just born in the location. What makes the outflow rate increase is that the average outflow rate for the local economy as a whole is a weighted average of the outflow rate of incoming and new born agents (p) and incumbent agents. Since the incumbents' outflow rate is zero, the average outflow rate increases after an increase in gross inflows. As a result, the gross inflow rate in period 2 exceeds the net inflow rate.

[31] Remember that the timing of the model is such that workers cannot, by assumption, move in from other locations to take advantage of the positive labor market conditions in period 1.

- The following periods are similar to period 2, with net flows moving back toward steady state, gross flows exceeding steady state, and gross inflows exceeding net inflows.

The response to a negative productivity shock can be of two types. First, if the shock is not too large, the impulse response is simply the negative of the one plotted in Figure 6, with net flows dropping, outflows remaining constant on impact and gross inflows tracking net flows. Notice that in this situation, as in the one above, badly matched agents account for the entire gross outflow from the location. If, however, the negative shock is large enough in absolute value, then gross inflows go to zero on impact. In this case the local economy reaches a corner solution and highly matched incumbent workers (i.e. those characterized by idiosyncratic shock υ_h) might choose to leave the location. This situation is represented in Figure 7. The figure shows a simulation of gross inflow and outflow rates over time, with inflows hitting zero in period 10. In this case, about nine percent of the highly matched agents choose to migrate out of the location, giving rise to an outburst of outflows. Notice that, except for these occasional spikes in outflows, the outflow series is smoother than the inflow series.

What moments does this adjustment path imply? Compare first the standard deviation of net inflow, gross inflow and gross outflow rates. On impact, following the positive labor demand shock, net inflows rise while gross outflows remain constant. Over time gross outflows respond to the higher inflow of workers. The fact that gross outflows are above steady state in the periods following the shock implies that gross inflows exceed net inflows. Thus, gross inflows are more volatile than net inflows, and the latter are more volatile than gross outflows. In terms of correlations, in the period following the shock all three flows tend to move together leading to a high correlation between any two of them. In the data only gross and net inflows are highly correlated. Gross inflows and outflows are positively correlated, but less than the amount predicted by the model. The latter predicts that net flows and gross outflows are positively correlated in the cross-section, while they tend to be slightly negatively correlated in the data. Finally, gross and net flows are highly positively correlated over time in the model and in the data.

Average Wages. Consider now the effect of a positive productivity shock occurring in period 1 on average local wages. The latter can be decomposed into the unit price of labor and the average efficiency units in the location:

$$\underbrace{\ln w_{jt}}_{\text{unit price of labor}} + \underbrace{\lambda_{jt} \log \upsilon_l + (1 - \lambda_{jt}) \log \upsilon_h}_{\text{average efficiency units}}. \tag{18}$$

First, the unit price of labor in period 1, w_{j1}, rises on impact, because of the unanticipated nature of the shock. In response to this shock, there will be a gross inflow of workers in the location. This inflow lowers the marginal product of labor and brings w_{jt} back to its steady state value starting from period 2. The unit price of labor is shown as the dashed line in Figure 8.

Second, the inflow of workers has also an effect on the average efficiency units in the location (the second term in equation (18)). The workers that flow into the location in period 1 (together with the ones born at the end of period 1) draw idiosyncratic shocks in period 2. The average idiosyncratic match drawn by this class of workers equals 1. In contrast, due to the selection associated with workers' moving decisions, all other workers are characterized, at beginning of period 2, by a high match υ_h with the location. It follows that a positive shock to the location's productivity in period 1 will induce an increase in the proportion λ_{j2} of low matches and therefore a fall in the average efficiency units of labor.[32] This second effect, represented by the dotted line in Figure 8, is however quantitatively very small. It follows that the impulse response of average wages tracks closely the impulse response of the unit price of labor.

7 The Value of Excess Workers' Flows

The estimated parameters can be used to compute the welfare effect of eliminating excess workers flows, i.e. the difference between gross flows and absolute net flows, in this economy. This exercise parallels Jovanovic and Moffit (1990)'s computation of the value of job-specific information accumulated by workers. In my model, excess flows of workers across states are due to workers' learning about their idiosyncratic productivity in different locations. To measure the contribution of this information to aggregate output, I consider a version of the model in which workers are all identical ex-ante and ex-post and characterized by one unit of efficiency. This is just the unconditional average of υ_l and υ_h. By construction, then, all matches between workers and locations are identical, and workers will only move from a location in order to earn a higher unit price of labor. The equilibrium in this alternative setting is easy to characterize. Notice, in fact, that there are no direct mobility costs in this economy. Thus, a worker will always choose to move away from a location that offers a lower expected unit price of labor than what can be obtained elsewhere. It follows that in the equilibrium of this alternative model expected unit prices of labor will be equalized across locations. Formally, an optimal allocation will be characterized by the following condition:

$$E\left[w\left(s'\right)|s\right] = (1-\beta\delta)\,e,$$

where the left-hand side is the expected price of labor in a location characterized by current state s, and the right-hand side is the flow value of moving to a different location offering expected utility e. Replacing in the expression above the new unit price of labor:

$$w\left(s'\right) = \tau z'\left(y'\right)^{\tau-1},$$

and solving for population y' yields:

$$y' = \left[\frac{\tau E\left[z'|z,\varepsilon\right]}{e\left(1-\beta\delta\right)}\right]^{\frac{1}{1-\tau}},$$

where $E\left[z'|z,\varepsilon\right]$ denotes the expectation of z' conditional on z and ε.

In order to obtain the equilibrium value of e, it is enough to replace this equation in the equilibrium condition (16). Aggregate output in the economy can then be easily shown to be equal to

$$\left[\int\int\left(E\left[z'|z,\varepsilon\right]\right)^{\frac{1}{1-\tau}}\Phi_{z,\varepsilon}\left(z,\varepsilon\right)dzd\varepsilon\right]^{-\tau}\int\int z^{\psi}\varepsilon^{\rho(1-\psi)}u^{1-\psi}E\left[z'|z,\varepsilon\right]^{\frac{\tau}{1-\tau}}\Phi_{z,\varepsilon}\left(z,\varepsilon\right)q\left(u\right)dzd\varepsilon du,$$

[32] It follows that the incoming workers will have, on average, lower wages than incumbents. This is consistent with the evidence of section (2).

where $\Phi_{z,\varepsilon}$ denotes the stationary density of (z, ε).

Notice that in the absence of mobility costs, aggregate output is the appropriate measure of welfare in this economy. This value can be compared to its counterpart in the model of section (3). Using the estimated parameters suggests that the contribution of excess workers flows to aggregate welfare is about 1.16 percent of aggregate output in the economy of section (3).

To understand what's behind this number, assume for simplicity that the non-negativity constraint on gross inflows never binds. This assumption, while literally not true, represents a good first approximation as, in practice, the constraint on gross inflows rarely binds. Under this circumstance, the percentage gain on aggregate output from excess flows can be shown to be equal to:

$$1 - \left[\frac{1-\delta p}{1-\delta p \upsilon_l}\right]^{\tau} > 0, \tag{19}$$

where the inequality is due to the fact that $\upsilon_l < 1$. Intuitively, the gain from gross flows is larger the lower the productivity υ_l of a bad match, the more often gross flows occur (i.e. the larger p), the longer an agent's lifespan, as indexed by δ, and the larger the income share τ of output. Notice that the key parameter in (19) is υ_l, with $\upsilon_l = 1$ (i.e. the unconditional expectation of υ) yielding zero gain from excess flows. The estimated value of υ_l is 0.966, which is close enough to one to lead to relatively small gains from excess migration flows.[33] Of course, given the relative simple structure of the model and the cross-sectional nature of the earnings data, this figure has to be interpreted cautiously.

8 Conclusions

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, with outbursts of outflows being followed by long periods of relatively constant flows.

The estimated parameters of the model are used to compute the gains from workers' excess reallocation across U.S. states. The latter represent about one percent of output.

I conclude with a few remarks about future research in this area. There are two features of the data that the model cannot reproduce. First, the model predicts that outflow rates in a location are highly positively correlated with both gross and net inflows. This is at odds with the data, as observed outflow rates display a small positive correlation with gross inflows and a small negative correlation with net inflows. One way to address this problem would be to introduce in the

[33] Plugging the estimated parameters in (19) yields a gain of 0.9 percent of aggregate output.

model a shock that affects outflows directly. For example, in the current version of the paper, if an agent draws an idiosyncratic shock υ_h, the good match with the location lasts until the agent dies or voluntarily leave the location. Alternatively, one could assume that the duration of all idiosyncratic matches in a location is stochastic. In this setting, locations hit by aggregate "destruction" shocks would experience a gross outflow of workers, which is likely to reduce the correlation between gross outflows and the other flows, in addition to making outflows more volatile.

Second, the model can reproduce only a fraction of the observed dispersion in average real earnings across states and the observed correlation between wages and net flows. The observed dispersion of average earnings can be attributed to two potentially complementary causes. On the one hand, population flows might be subject to adjustment costs that slow down the process of β-convergence of earnings across locations. This will result into larger cross-sectional dispersion at a point in time. On the other hand, differences in real earnings might capture differences in amenities across locations. As discussed above, it is not easy to determine a-priori which amenities people value. The positive correlation between real earnings and measures of "good weather" across states suggests that weather might not be the appropriate amenity to consider. Of course, it could be that the observed differences in real earnings are due to measurement error because of the difficulty of accurately measuring the level of prices in different locations. The availability of better data is crucial for further progress in this area.

More generally, the model introduced here can be naturally extended to evaluate the impact of international migration (i.e. net inflows of population from outside the aggregate economy) on the wages and internal mobility of natives. Borjas et al. (1997) carry out such exercise using a reduced-form model of internal net migration flows. Extending their analysis to consider gross flows would shed light on the process by which states absorb larger external migration. Does the latter lead to larger outflows of natives or less inflows from other states? Addressing this and related questions is an interesting topic of future research.

References

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.

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

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

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

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.

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.

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

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

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

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

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

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

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

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

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.

A Data Appendix

Sample Selection

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

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

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

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

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

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

I also denote by $\overline{y}_{j\tau}$ the total population that was interviewed in year τ'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 τ 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 τ 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 τ is defined as the difference between inflow and outflow rates:

$$\widehat{d}_{j\tau} = \widehat{x}_{j\tau} - \widehat{o}_{j\tau}.$$

Accounting for Heterogeneity

There are (at least) two sources of heterogeneity I need to worry about. The first concerns heterogeneity among locations (U.S. states) in the demographic composition of their population. For example, if state A has a younger workforce than state B, and, if younger individuals are characterized by higher mobility, then measures of outflows from state A will be higher than from state B due to these demographic differences. Given that, in the model, all individuals have the same age I need to control for these differences when constructing measures of outflows and inflows. The second, equally important, issue is that while in the model all locations are identical in terms of their "physical attributes", in the data, instead, different states are characterized by very different geographic features, sizes, etc. For example, outflow rates might be higher from smaller (size being measured in terms of land area) states.

In practice, adjusting the data to take into account these concerns is not straightforward, because while some of the heterogeneity across states can be safely taken as exogenous (e.g., their land area), other features such as their demographic composition are likely not to be so.

In what follows I proceed in two steps. First, in order to construct measures of population flows that are free from potential composition effects, I divide the population in each location (state) in different cells, defined by the following characteristics in the Census year:

- 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 (`ind`), including the unemployment state; 14 industries: (1) unemployed, (2) agriculture, fishing, forestry and mining, (3) utilities, (4) construction, (5) manufacturing, (6) wholesale and retail sales, (7) transportation and wharehousing, (8) information and communication, (9) finance, insurance, real estate and leasing, (10) professional, scientific, management, (11) educational, health, social service, (12) arts, entertainment, recreation, (13) other services, (14) public administration.

Denote each cell by g and the collection of cells by G. There are 490 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:

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

Similarly, it is possible to define net flows.
The second step of the procedure consists of controlling for geographical and historical differences among U.S. states and the effect these might have on population flows. I do this by running separate cross-sectional regressions of inflow and outflow rates on the following state-level variables: (1) land area, (2) year when state joined the U.S., (3) number of metropolitan areas within a state with population larger than 500,000 (computed including all inhabitants) in the 2000 Census. The inflow and outflow rates presented in the main body of the paper are the residuals from these regressions.

Real Weekly Earnings

Data on workers' weekly earnings were computed from the Census 2000 by summing annual wage income (`incwage`), business and farm income (`incbus00`), and welfare income (`incwelfr`), and dividing the sum by the number of weeks worked (`wkswork1`). Each source of income refers to the year 1999. 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 for 1999. In a few instances reported earnings by self-employed individuals were negative, and these observations have been dropped. Given that earnings refer to 1999 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 1999. I have also dropped these individuals from the sample.

The earnings data were deflated using the ACCRA Cost of Living index for the third quarter of 1999. This index measures relative price levels (gross of taxes) for consumer goods and services (including housing) in a number of U.S. cities.[34] This number varies from quarter to quarter, and the third quarter of 1999 was selected to maximize coverage of locations (330). Using information on the workers' metropolitan statistical area or PUMA of residence in 2000 (found in the Census), I have matched workers with a value of the Cost of Living index in their area of residence. Unfortunately, the limitations in the coverage of the ACCRA index prevent one from being able to do so for all workers in the Census sample. In particular, only 53 percent of the workers could be matched with a value of the Cost of Living Index. The results reported in the main text of the paper refer only to these workers (about 2.5 million individuals).

The logarithm of real weekly earnings was regressed on the following variables: 48 dummies for workers' state of residence in 2000 (`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", from `raced`), 14 sectoral dummies (from `ind`), and 26 occupational dummies (from `occ`). The R^2 of this regression was 30 percent.

March Current Population Survey

March CPS data for the years 1999-2003 were used to compute the fraction of individuals moving across U.S. states for job-related reasons. During these years the March CPS contains a question regarding an individual's primary reason for changing residence with respect to the previous year. The Census does not contain such question. The March CPS questionnaire identifies 16 different primary reasons for moving, with "New job or job transfer" (33 percent of the answers) representing the most-frequent single answer, followed by "Other family reasons" (13 percent). In order to compute the fraction of individuals moving for job related reasons I have first applied the sample selection criteria listed above to the March CPS data. Then, the survey's different 16 reasons for moving were aggregated into two categories: job-related and non job-related reasons. The aggregation is relatively straightforward, with the exception of moves motivated by the desire of new, better or cheaper housing. I have included those reasons in the job-related move. The rationale

[34] Given that the index only measures relative price levels, it cannot be used to compare the price level in the same location over time.

for this choice is that the wage data used to calibrate the model have been deflated using price indices that include housing prices.

B 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 e and θ, one for finding the equilibrium of the model for given θ, and one for finding θ in order to match the empirical moments of interest. Every change in θ entails a new equilibrium e, while a new e requires the computation of the associated value function.

Step 1 (Guess). Start from an initial guess for the parameter vector θ and for the value of migration e. The guess for e is updated in Step 3 below, while the guess for θ 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 $s = (y, n_{-1}, z, \varepsilon)$) and because the procedure involves numerical integration of the value function with respect to the density $q(u)$ 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 $x(s)$ and the probability of outflow $q(s)$ for an agent with match υ_h as functions of the state vector s. These two functions allow one to recover all the other variables of interest, in particular the location's population $y(s)$ conditional on s.

Step 3 (Equilibrium). Solve for the equilibrium value of e by defining the function

$$f(e) = \int y(s)\Phi\left(s\right) ds - 1. \tag{20}$$

The value e^* such that $f(e^*) = 0$ represents the equilibrium value of migration. The integral in equation (20) is computed by simulating the economy for a very large number of periods (5 million), obtaining $\{y_t\}_{t=1}^{T}$ and approximating the integral in (20) with the sample average:

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

In practice the zero of (20) is computed using a simple bisection procedure. The function $f(e)$ is decreasing in e because a higher value of migration must be associated with higher expected wages which reduce firms' demand for labor. Notice that for each candidate value of e it is necessary to go back to Step 2 and solve the dynamic programming problem again.

Step 4 (Estimation). Given e^*, it is feasible to compute the equilibrium value of all the variables of interest. The vector θ is estimated by constructing the model counterpart of the six moments listed in the text (section 5) and choosing θ 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 simulating for $S = 100$ times artificial data for $J = 48$ locations for a number of $P = 1,000$ periods. For each simulation $s = 1, 2, .., S$, the data for all but the last 12 periods are then discarded. The data for the last of the 12 periods are instead used to compute the cross-sectional moments, while the whole 12 periods are then used to compute the time-series moments. Recall, though, that each period represents 5 years, while the correlations of flows across Census years refer to 10 years, therefore in practice only 6 time-series observations for each location and simulation are used. Six moments are computed for each of the S simulations. Each moment is then averaged across simulations and 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 θ 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, θ 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 θ 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 smaller than 10^{-4} in absolute value.

Figure 1 - Inflow and Outflow Rates

Correlation Coefficient (Raw Data): 0.66

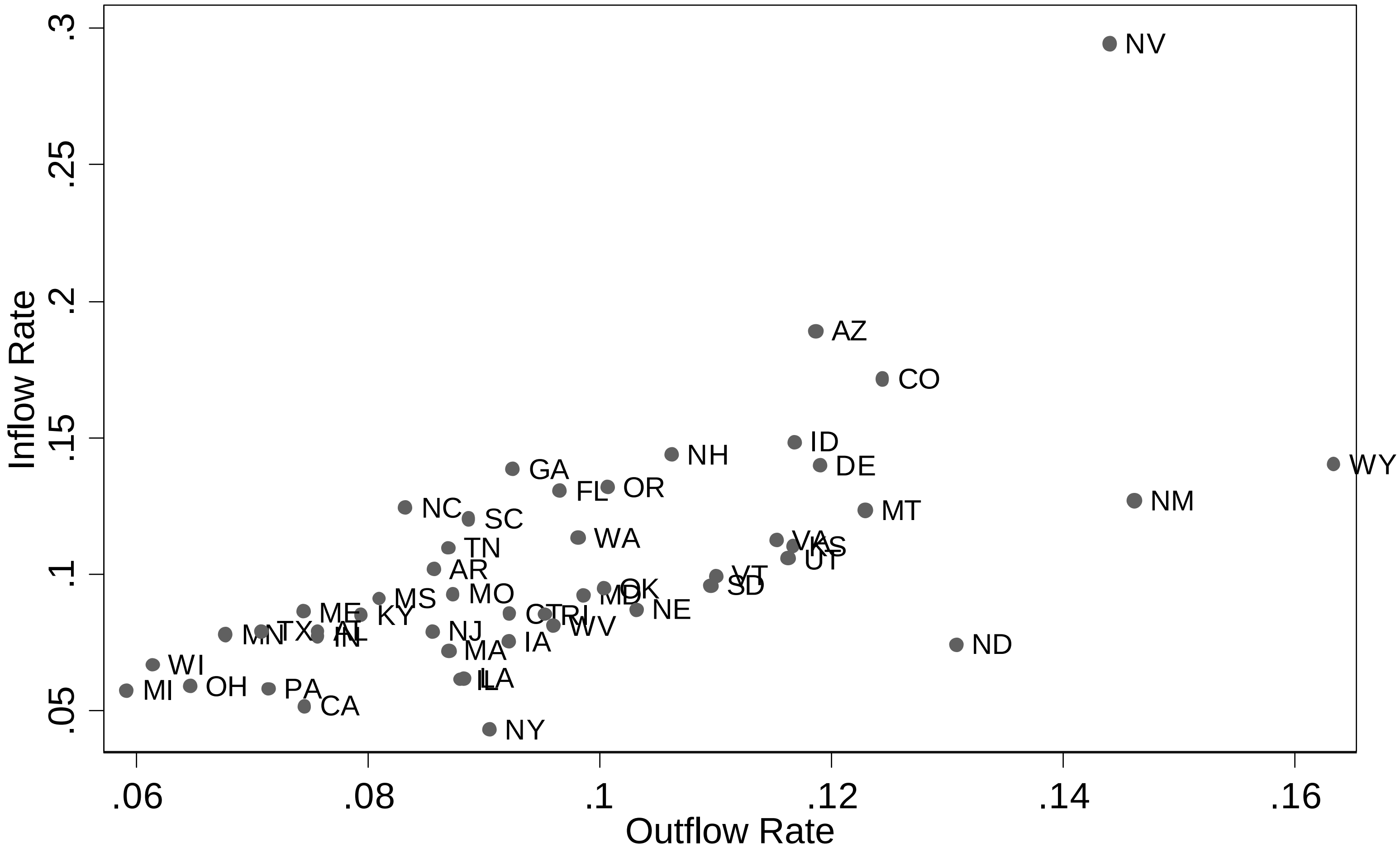


Figure 2 - Inflow and Net Flow Rates

Correlation Coefficient (Raw Data): 0.85

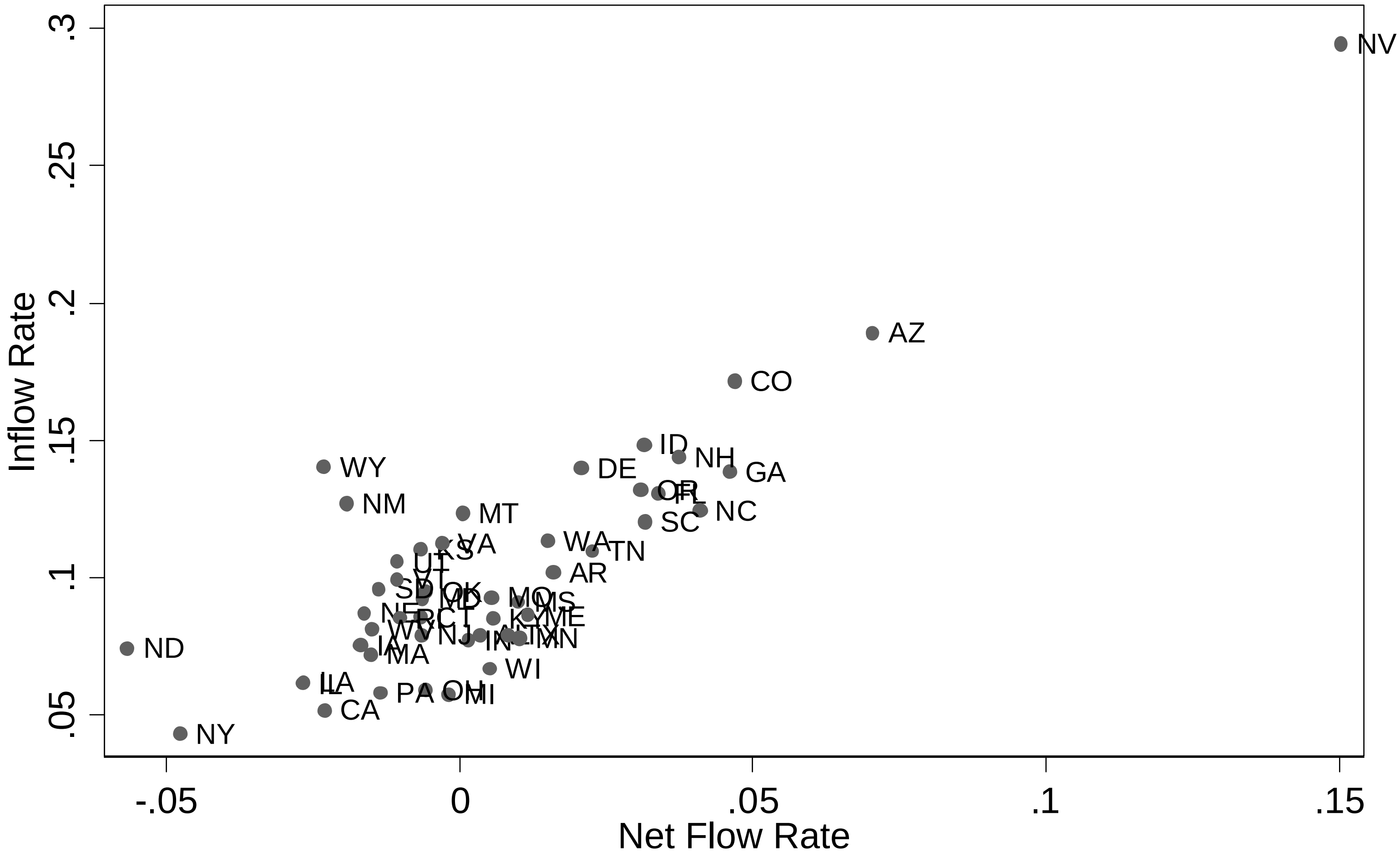


Figure 3 - Net Flow and Outflow Rates

Correlation Coefficient (Raw Data): 0.16

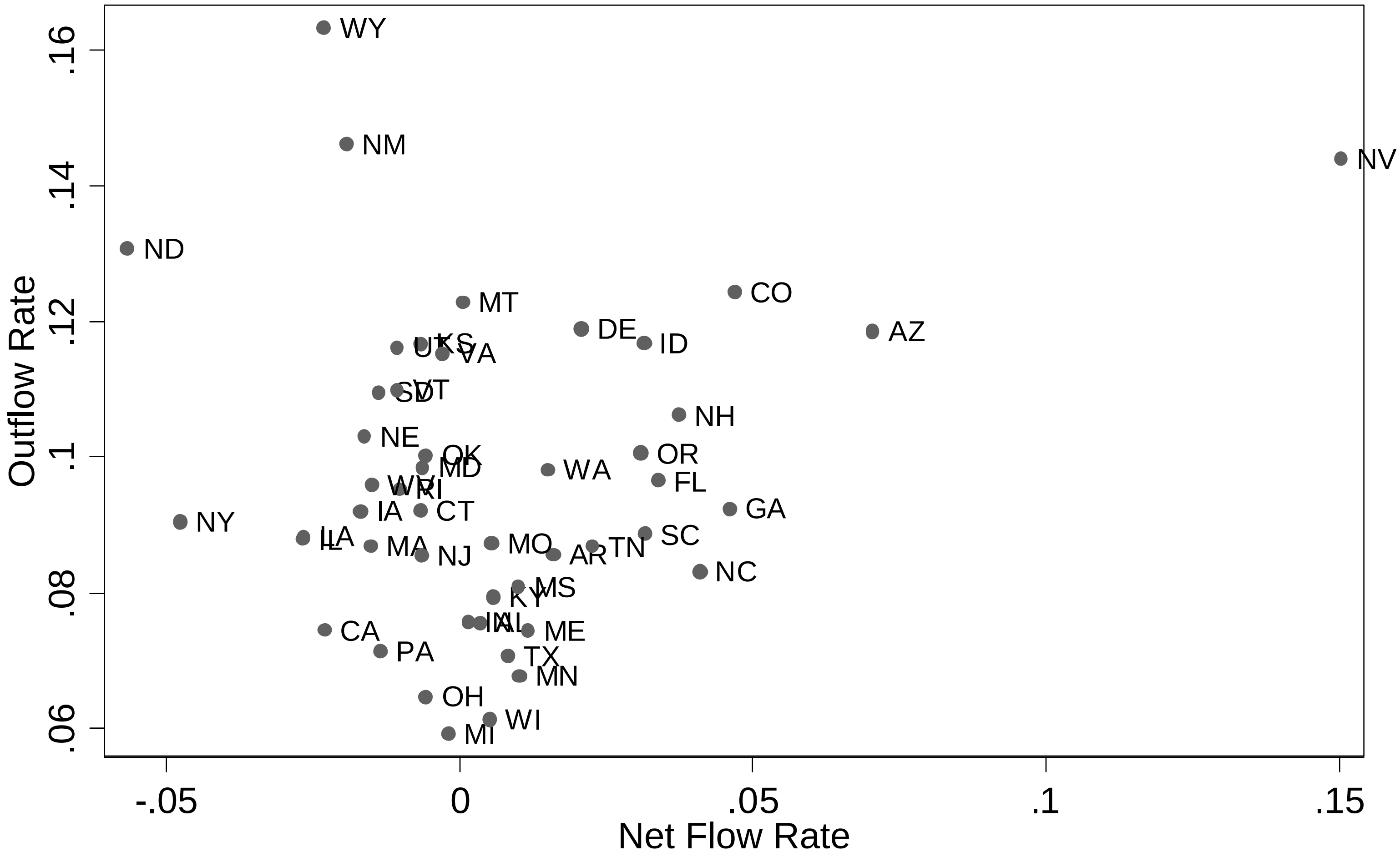


Figure 4 - Net Flow Rates and Average Earnings

Correlation Coefficient: 0.43

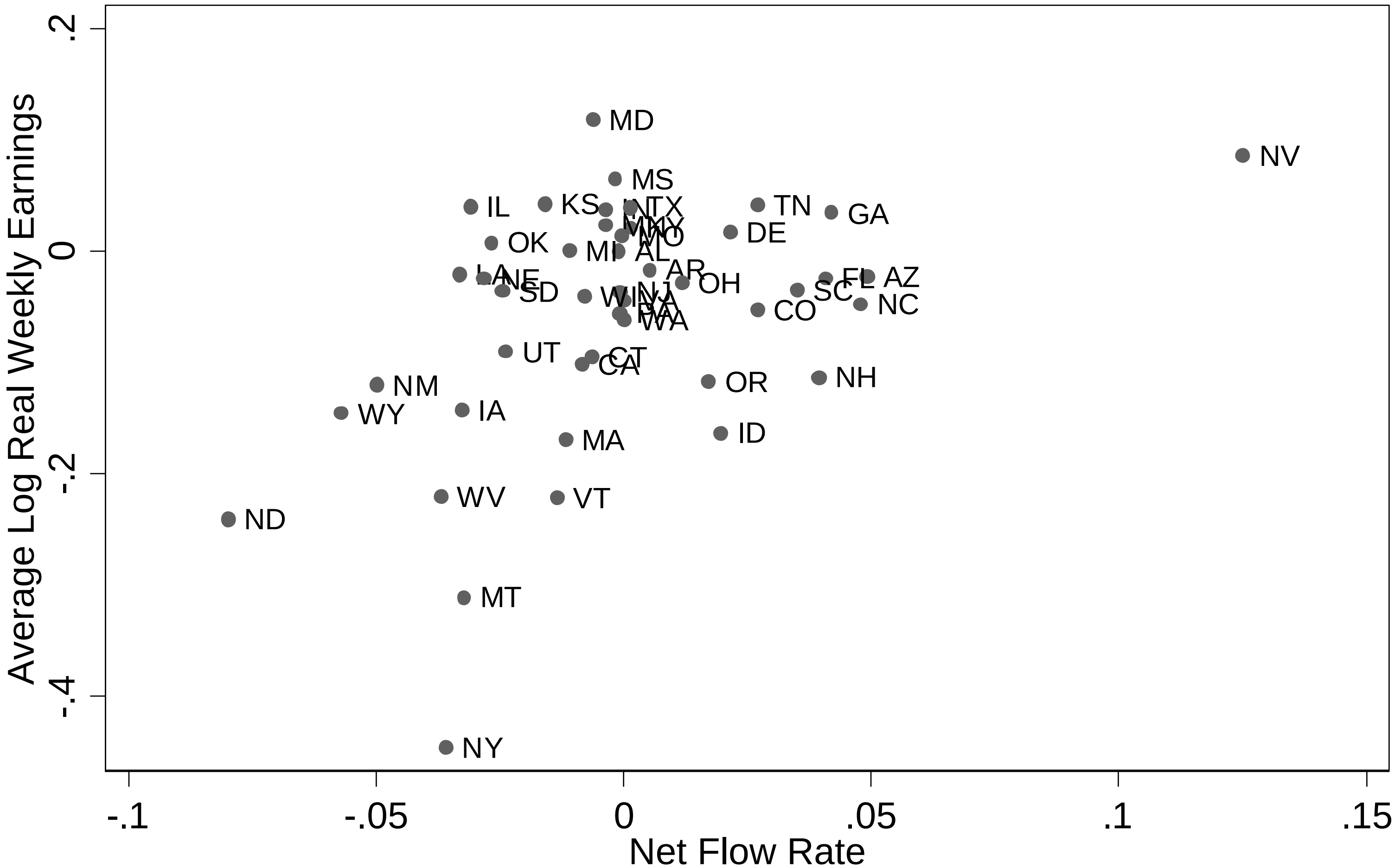


Figure 5 - Net Flow Rates and Hot Weather

Correlation Coefficient: 0.44

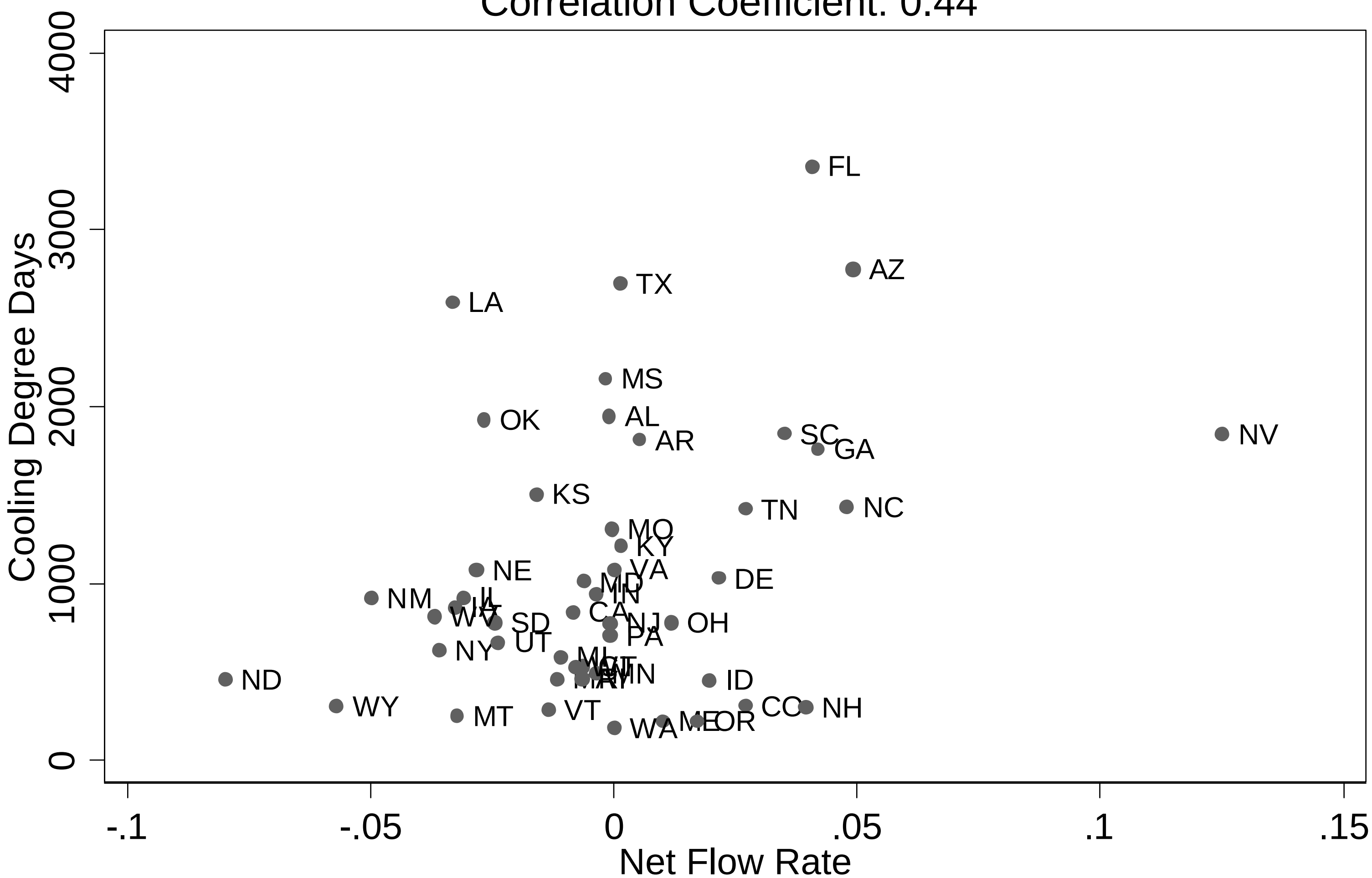


Figure 6: Impulse Responses of Gross and Net Flows to Local Labor Demand Shock

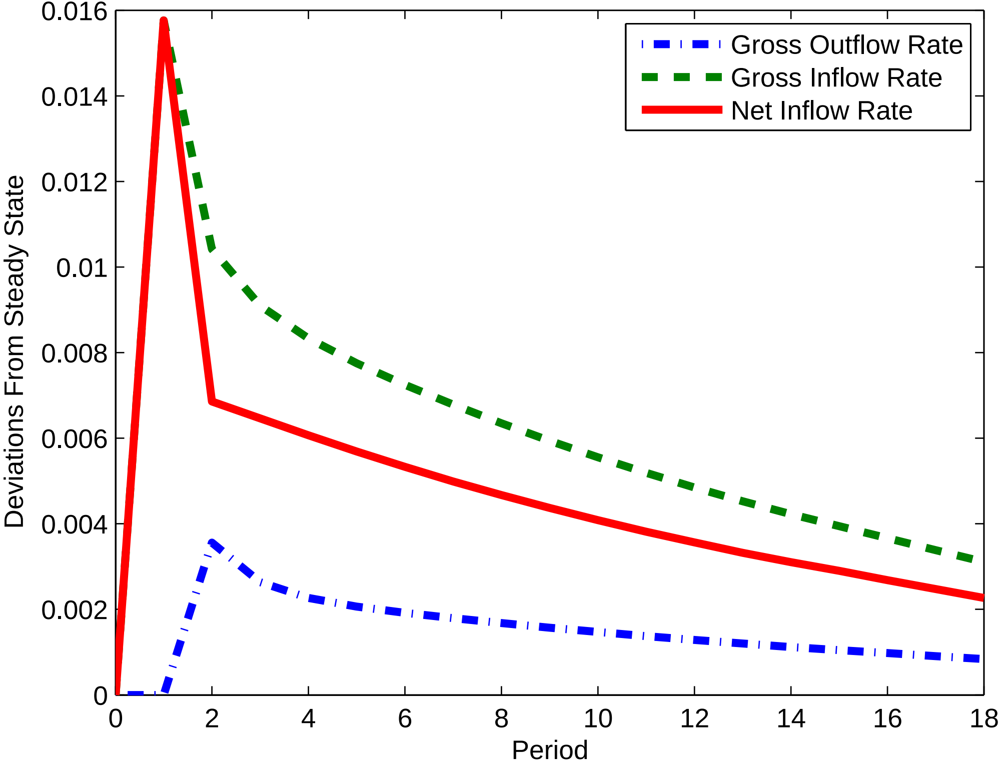
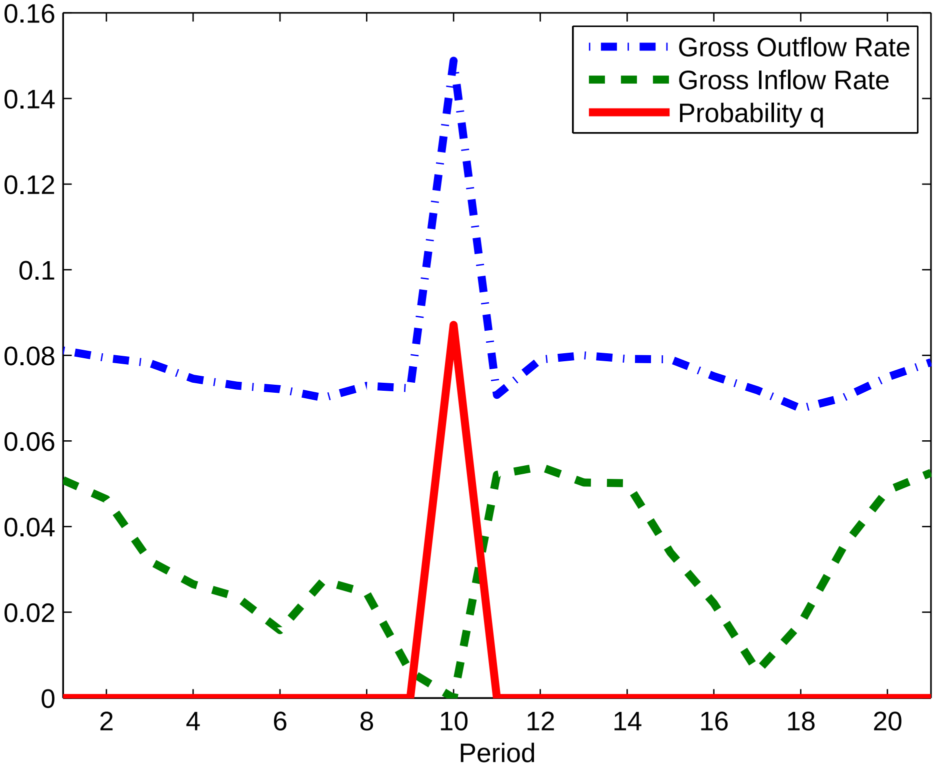


Figure 7: Time−Series of Outflows and Inflows with Binding Constraint on Gross Inflows



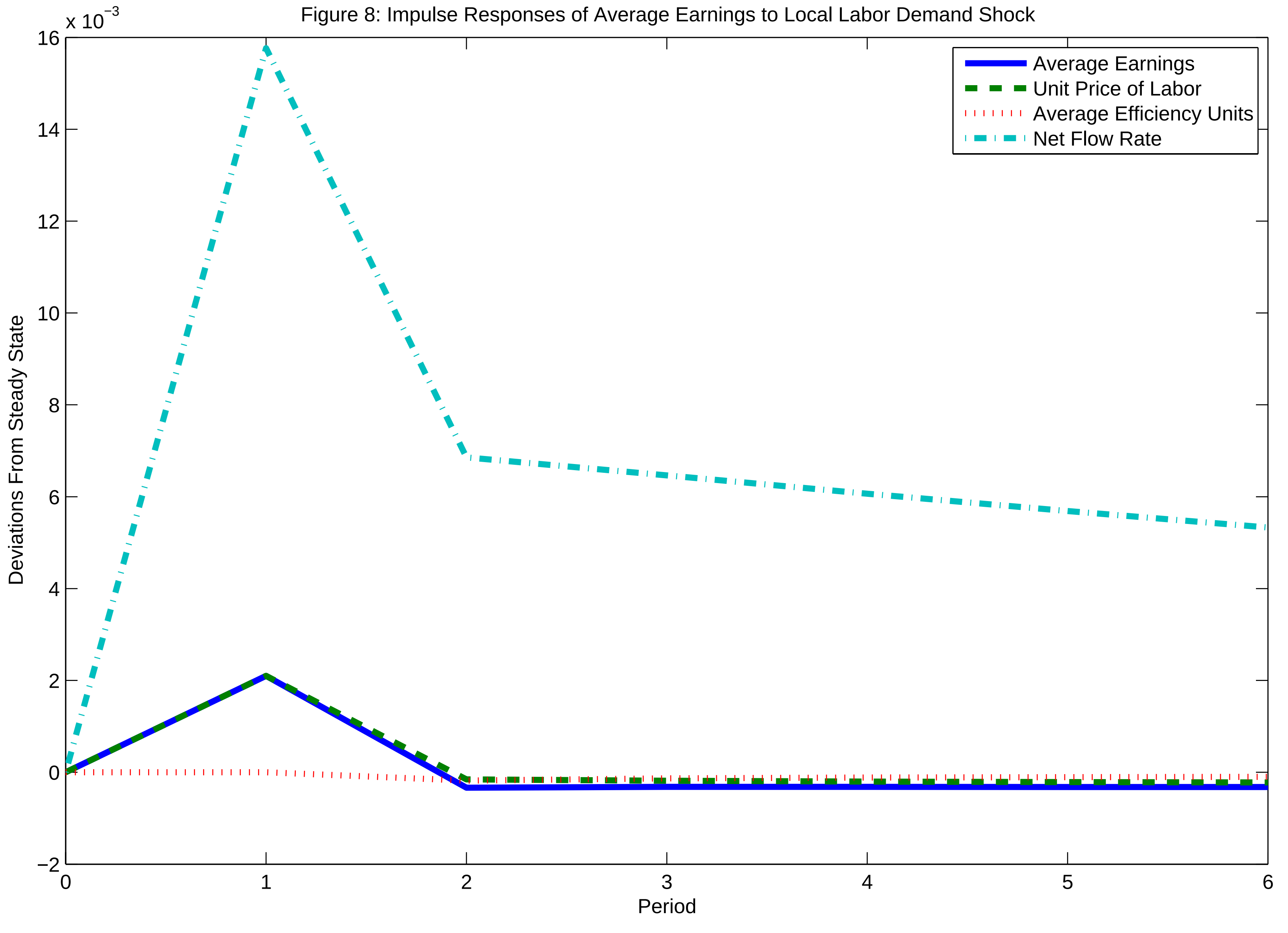


Figure 8: Impulse Responses of Average Earnings to Local Labor Demand Shock
x 10^−3
16
14
12
10
8
6
4
2
0
−2
0
1
2
3
4
5
6
Period
Deviations From Steady State
Average Earnings
Unit Price of Labor
Average Efficiency Units
Net Flow Rate