Re-estimation of the U.S. Inter-State Economic Migration

April 2017

1. Introduction

In the past decade, the U.S. economy has experienced the financial crisis of 2008 and the worst recession since the Great Depression. An interesting observation of the U.S. inter-state migration is that the trend of total number of inter-state migrants has been following the U.S. economy conditions. The total inter-state migration has been increasing since 2000 and reached its peak in 2006. Before the recession, many businesses started to lower costs by laying off workers during 2006-2007, and the inter-state migrants also started to drop. When the recession officially ended in 2009, the inter-state migration continued to decrease because of the lagged response of the labor market. The inter-state migration started to increase in 2011 when Americans' confidence in labor market finally picked up. Then it stayed at a fairly stable level, while the employment rises in recent years. It is reasonable to presume that the inter-regional economic migration pattern has changed after the recession.

Annual Resident Population Data from the U.S. Census Bureau shows changes in both the volume of inter-state migration and the origins and destination of migrants over the periods 2001-2008 and 2009-2014. Figure 1 presents the average annual net inter-state migration rate per 1000 population during 2001-2008. The blues states experienced net inflows of inter-state migrants, and the red states are net exporters of migrants to other states. The darker colored states experienced higher volume of net migration, and vice versa. The map shows that the population generally move from the mid-western and the northeastern states to the western and southern states, with a few exceptions. The large out-migrants from Louisiana is attributed to Hurricane Katrina. Among the states with net out-migrants, New York had the highest average annual migration rate of -9.995 per 1000 population; Nevada had the highest average in-migration rate of 19.976 per 1000 population.

Figure 1. Average Annual Net Inter-State Migration Rate per 1000 population, 2001-2008

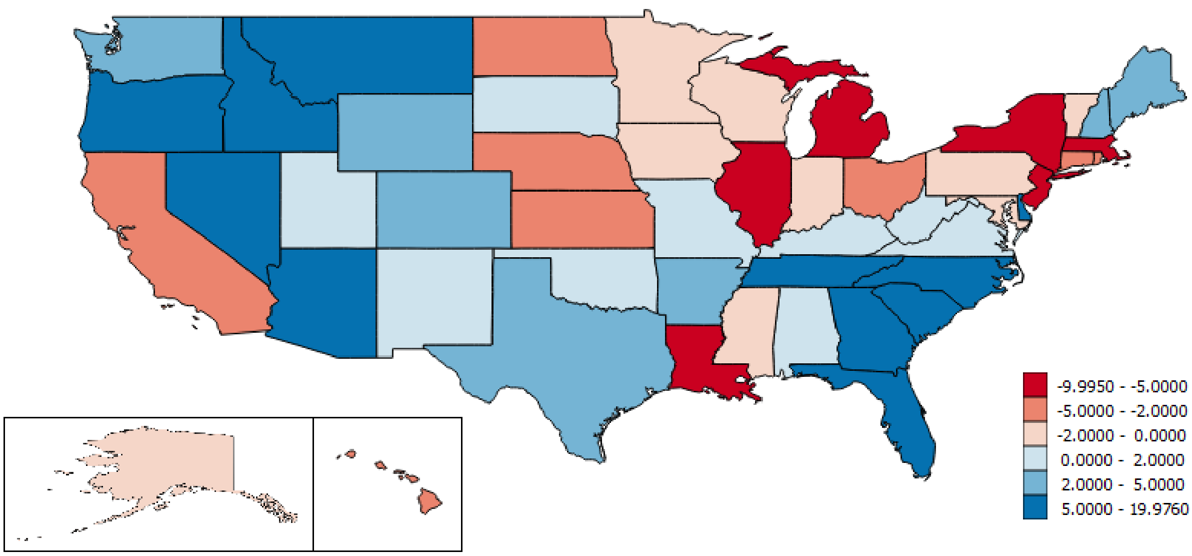
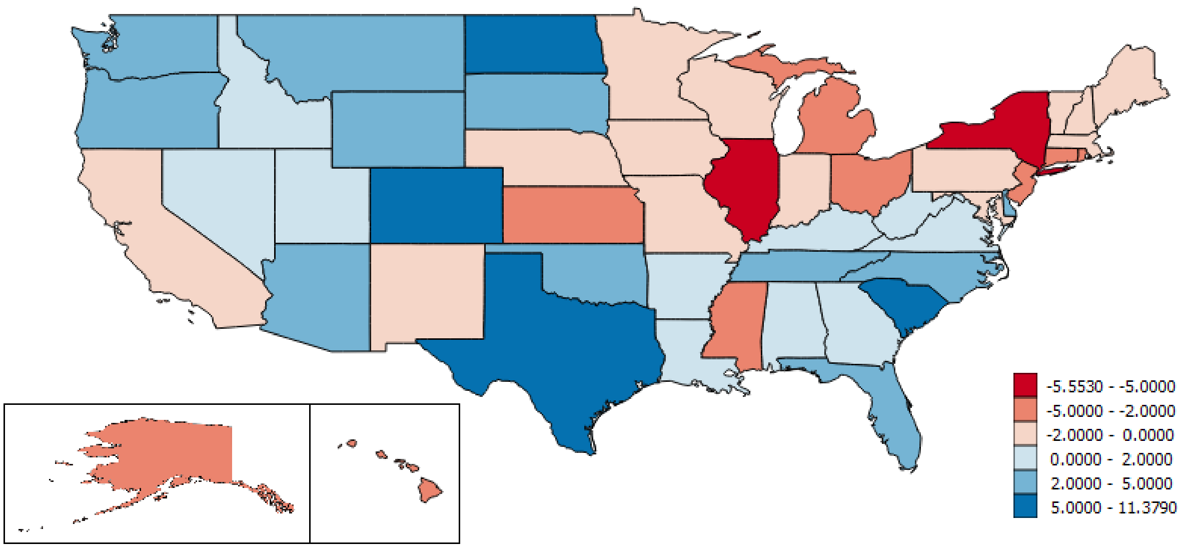


Figure 2 shows the average annual net inter-state migration rate per 1000 population during 2009-2014. Similarly, the blue states are states with net in-migration states and the red ones are the out-migration states. Compared to 2001-2008 period, the U.S. become less migratory across states in general, as indicated by the lighter color of states in the map below. The direction of net migration flows in most states have been stable over the two time periods. The northeastern and the mid-western states are still the major origins of domestic migration; while the western and the southern states are the major destinations. Two migrant-exporting states, North Dakota and Louisiana, have become migrant-importing states; four states, New Mexico, Missouri, New

Hampshire and Maine, experienced the change in opposite direction. New York remains to be the largest net out-migrating state, but the average net migration rate dropped from -9.995 to -5.553 per 1000 population. North Dakota, which used to be an out-migration state during 2001-2008, have experienced the highest net in-migration rate at 11.379 during 2009-2014, due to the new jobs created as a result of the oil boom.

Figure 2. Average Annual Net Inter-State Migration Rate per 1000 population, 2009-2014



The current migration equation parameters were last estimated using 2001-2008 data. This study revisits the economic migration equation with update data to examine changes in the inter-state migration patterns as people seeking opportunities for economic advancement and better life quality.

2. Estimating the REMI Migration Equation

2.1 The Model

The REMI migration equation assumes that the net inter-regional economic migrants are determined by changes in the relative employment opportunity REO_t^k, the relative compensation rate RWR_t^k, and the regional fixed amenity term λ^k, as shown in Eq (1).

$$\frac{ECMG_t^k}{LF_{t-1}^k} = \beta_1 \ln\left(REO_t^k\right) + \beta_2 \ln\left(RWR_t^k\right) + \lambda^k + u \tag{1}$$

where

$ECMG_t^k = Net\ economic\ migrants\ to\ region\ k\ in\ time\ t$;

$LF_{t-1}^k = Labor\ force\ in\ region\ k\ and\ time\ t-1$;

$REO_t^k = Relative\ employment\ opportunity$;

$RWR_t^k = Relative\ compensation\ rate$;

$\lambda^k = Fixed\ amenity\ term\ of\ region\ k$;

$\beta_1, \beta_2 = the\ coefficients\ to\ be\ estimated$;

$u = the\ error\ term$.

The relative employment opportunity is defined as the regional residence adjusted employment RAE_t^k as share of the regional labor force LF_t^k relative to its national average. $REO_t^k > 1$ indicates more employment opportunities in the region compared to the national average level; and $REO_t^k > 1$ indicates relatively less job opportunities in the region. Employment opportunity, as a most important driving force of migration, is expected to have positive impacts on net inflows of migrants. Eq (2) is used to calculate REO_t^k.

$$REO_t^k = \frac{RAE_t^k / LF_t^k}{RAE_t^u / LF_t^u} \tag{2}$$

where

$RAE = residence\ adjusted\ employment$;

k denotes region k; and u denotes the U.S.

The relative wage rate is the product of the relative industrial mixed compensation rate and the real disposable income as share of personal income relative to its national average, as defined in Eq (3).

$$RWR_t^k = \frac{CR_t^k}{CR_t^u} \times \frac{RYD_t^k / YP_t^k}{RYD_t^u / YP_t^u} \tag{3}$$

where

$CR_t^k = industrial\ mixed\ compensation\ rate\ in\ region\ k\ and\ time\ t;$

$CR_t^u = national\ average\ industrial\ mixed\ compensation\ rate\ weighted\ by\ region\ k's$
$industrial\ employment\ share\ at\ time\ t;$

$RYD_t^k = real\ disposable\ income\ in\ region\ k\ at\ time\ t;$

$RYD_t^u = real\ disposable\ income\ of\ the\ U.S. at\ time\ t;$

$YP_t^k = personal\ income\ in\ region\ k\ at\ time\ t;$

$YP_t^u = personal\ income\ in\ the\ U.S. at\ time\ t.$

The industrial mixed compensation rate CR_t^k is the sum of the regional compensation rate of each industry weighted by the employment share of the industry in the region, and it is generated by Eq (4):

$$CR_t^k = \sum_{i=1}^{n} \frac{E_{i,t}^k}{TE_t^k} \times C_{i,t}^k \tag{4}$$

where

$E_{i,t}^k = employment\ of\ industry\ i\ in\ region\ k\ and\ time\ t;$

$TE_t^k = total\ employment\ of\ in\ region\ k\ and\ time\ t;$

$C_{i,t}^k = compensation\ rate\ of\ industry\ i\ in\ region\ k\ and\ time\ t.$

Eq (5) defines the national average industrial mixed compensation rate, which is the sum of the national average compensation rate in each industry weighted by its employment share in region *k*, such that

$$CR_t^u = \sum_{i=1}^{n} \frac{E_{i,t}^k}{TE_t^k} \times C_{i,t}^u \tag{5}$$

where

$C_{i,t}^u = national\ average\ compensation\ rate\ of\ industry\ i\ and\ time\ t.$

Lastly, the regional fixed amenity term λ^k captures the unobserved factors in each region that reflect the life quality and affect people's migration decisions, such as the climate, environment, community safety, availability of education resources and health facilities, etc. These amenity factors are hard to measure but important in driving population redistribution among states, so they are included in the migration equation as regional fixed dummy variables. A positive amenity value indicates higher level of attractiveness of the region to economic migrants relative to the national average; a negative amenity value indicates relatively lower level of attractiveness in the region. β_1 and β_2 are the coefficients of interest that is to be estimated.

2.2 Methodology

In estimation of the migration equation, a major problem is the endogeneity of the two explanatory variables *REO* and *RWR*. Endogeneity can arise from omitted variables or reverse causality effects from the independent variables. Since economic migrants are pursuing higher quality of life when moving from one region to another, amenity factors (such as the climate, recreation facilities, cultural choices, etc.) could be important in their decision making process but difficult to measure. Some of these amenity factors that are important drivers of both the dependent and independent variables could be left out of the model. When the model compensates the omitted factors by overestimating or underestimating the effects of the included independent variables, the estimated coefficients of *REO* and *RWR* will be biased.

Another possible cause of endogeneity is the loop of causality between the net economic migrants and the independent variables. A region with more employment opportunities and higher wage rate attracts economic migrants to the region. Because the majority of economic migrants are working age people, the net inflows of migrants will not only contribute to the local human capital resources, but also stimulate the development of local business through an increase in demand, which will in turn generate more job opportunities and affect the average wage rate. In this context, OLS estimates will be biased even if additional controls are added to the model.

Econometric tests of the data also show that the independent variables are correlated with the regression error term. The dependent variable has causality effects on both *REO* and *RWR* at the 1% significance level. OLS regression assumes all explanatory variables are exogenous. In the presence of endogeneity, OLS regressions will produce inconsistent estimates and are reliable in predicting the economic migration.

To correct the endogeneity problem, the instrumental variable (IV) approach is used to estimate the migration equation. The migration equation can be formulated as a model system with Eq (6)-(8). Because the explanatory variables *REO* and *RWR* in Eq (6) are the dependent variables of other equations in the system, the error terms among the equations are expected to be correlated. Three-stage least squares (3SLS) regression is utilized to apply the IV approach. The model uses generalized lease squares (GLS) to account for the correlation structure in the disturbances across the equations.

$$\frac{ECMG_t^k}{LF_{t-1}^k} = \beta_1 \ln(\widehat{REO}_t^k) + \beta_2 \ln(\widehat{RWR}_t^k) + \lambda^k + u \tag{6}$$

$$\ln\left(REO_t^k\right) = \alpha_1 EMX_t^k + \alpha_2 WMX_t^k + \alpha_3 FM_t^k + \alpha_4 fuel_t^k + \delta^k + \varepsilon_1 \tag{7}$$

$$\ln\left(RWR_t^k\right) = \gamma_1 EMX_t^k + \gamma_2 WMX_t^k + \gamma_3 FM_t^k + \gamma_4 fuel_t^k + \mu^k + \varepsilon_2 \tag{8}$$

where

$\ln(\widehat{REO}_t^k) = the\ predicted\ value\ of \ln(REO_t^k)$;

$\ln(\widehat{RWR}_t^k) = the\ predicted\ value\ of\ \ln(RWR_t^k)$;

$EMX_t^k = industrial\ mixed\ employment\ growth\ rate\ in\ region\ k\ at\ time\ t$;

$WMX_t^k = industrial\ mixed\ wage\ rate\ in\ region\ k\ at\ time\ t$;

$FM_t^k = federal\ military\ employment\ as\ share\ of\ labor\ force\ in\ region\ k\ at\ time\ t$;

$fuel_t^k = the\ relative\ fuel\ cost\ in\ region\ k\ at\ time\ t$;

$\delta^k, \mu^k = the\ regional\ fixed\ effects\ in\ the\ first\ stage\ regressions$;

$\varepsilon_1, \varepsilon_2 = the\ error\ terms\ in\ the\ first\ stage\ regressions.$

In the first stage, the predicted values of *REO* and *RWR* are developed with a set of instrumental variables and the regional fixed effects through regressions. This stage is critical for the consistency of parameter estimates. In the second stage, a consistent estimate for the covariance matrix of the equation disturbances is obtained. In the third stage, the predicted values of *REO* and *RWR* are used in place of the actual values of them with the covariance matrix to perform a GLS estimation.

Credibility of the instrumental variable method largely depends on the selection of suitable instrumental variables, since a weak instrument will result in severe loss of precision of the IV estimates and may bias the estimates in the same direction of OLS. A valid instrumental variable must have a significant effect on the endogenous explanatory variables, but cannot be correlated with the error term in the structural equation.

The relative employment opportunity *REO* represents the possibility of being employed in the region. Since the economic migrants are driven by possibility of economic advancement, many variables that have effects on the *REO* also have direct effects on the dependent variable. In order to make a reliable prediction of *REO*, an industrial mixed employment growth rate variable *EMX* is constructed using data from the 23-sector PI+ model. *EMX* is defined as the sum of the national average employment growth rate of each industry weighted by the industry's employment share in the region, such that

$$EMX_t^k = \sum_{i=1}^{n} \frac{E_{i,0}^k}{TE_0^k} \times g_{i,t}^u \quad (9)$$

where

$E_{i,0}^k = \textit{the employment of industry } i \textit{ in region } k \textit{ at the initial time period};$

$TE_0^k = \textit{the total employment in region } k \textit{ at the initial time period};$

$g_{i,t}^u = \textit{national average employment growth rate of industry } i \textit{ in time } t.$

EMX gives the hypothetical employment growth rate in each region assuming that each of the industries grow at the same national average rate. Using the national employment growth rate by industry guarantees that the instrumental variable does not directly affect inter-state economic migration. Meanwhile, the regional industrial structure captured by the employment weights determines the possibility of employment in the region. A state with a larger share of nationally fast growing industries will provide more employment opportunities; a state dominated by declining industries will have relatively less employment opportunities. Besides *EMX*, another instrument variable used to predict *REO* is the federal military (*FM*). The military employment have significant effects on regional employment opportunity, but have no direct influence on economic migrants.

Similarly, industrial mixed wage rate *WMX* is created as an instrument to predict the relative compensation rate *RWR*. Eq (10) defines *WMX* as the national wage rate weighted by the initial-year's employment share of each industry summed across industries. *WMX* is the hypothetical wage rate assuming each industry in the region pays the national average rate. Since the regional differences in wage rate is eliminated, the net economic migration to the region will not be directly influenced by *WMX*. The regional relative wage rate will be higher if the region has a larger share of industries with higher national average wage rate; otherwise the regional relative wage rate will be lower.

$$WMX_t^k = \sum_{i=1}^{n} \frac{E_{i,0}^k}{TE_0^k} \times W_{i,t}^u \tag{10}$$

where

$W_{i,t}^u = national\ average\ wage\ rate\ of\ industry\ i\ in\ time\ t.$

In addition to WMX, another instrument used to predict RWR is *fuel*, which is the fuel cost of the region relative to the national average. Fuel price directly affects costs such as transportation, heating and manufacturing, which in turn affect the prices of variety of goods and services. A higher fuel price usually means higher cost of living, which requires a higher compensation rate to afford living in the region.

Fixed effects are also applied to the first stage regressions to include the unobserved characteristics of the region that would affect the regional employment opportunity and compensation rate. As mentioned, these unobserved factors could include environment, public safety, accessibility to various resources in favor of business development, etc.

2.3 Data

The REMI migration equation was last estimated using 2001-2008 state-level data from PI+ V1.2 model. According to the Census Bureau Annual Resident Population Estimate, the rate of net domestic migrants[1] had dropped dramatically during the recession, and then stayed at a low level after the recession. The recession has a substantial impact on people's confidence in the job market and the dynamics of domestic migration. The question is: how does it change the causes and extent of the movement of domestic economic migrants? This study uses 2009-2014 state-level panel data generated from PI+ V2.0 model to re-estimate how the economic migrants respond to changes in the regional economic factors after the recession and how the relationship between regional amenities and migrants is distinct through time.

2.4 Results

Using on 2009-2014 state-level panel data from PI+, the parameters of the migration equation are estimated by instrumental variable regression with fixed effects. *EMX, WMX, FM,* and *fuel* are the instrumental variables employed in the first stage regressions to predict *REO* and *RWR*. The estimated coefficients are shown in Table 1:

Table 1. Estimates of Migration Equation

Variable	Estimated Coefficient	Std. Error	p>|t|
ln(REO)	0.4555	0.0890	0.000
ln(RWR)	0.2709	0.1060	0.011

[1] The proportion of people who moved to other states from their home states during the survey year.

Both of the regional employment opportunity and the regional relative compensation rate have significant positive impacts on the regional net economic migration rate, which is consistent with the expectation. A 1% increase in regional relative employment opportunity will cause the net economic migration rate to increase 0.4555%; and a 1% raise in regional relative compensation rate will increase the net economic migration rate by 0.2709%. By contrast, the corresponding parameters from OLS are 0.1557 for *ln(REO)* and 0.0015 for *ln(RWR)*, which are significantly under-estimated.

I test the endogeneity of *REO* and *RWR* with the Durbin and Wu-Hausman tests. The null hypothesis is that the variables being tested can be treated as exogenous. The Durbin test returns a Chi-squared statistic of 6.1053 with a p-value of 0.0472; and the Wu-Hausman test returns an F-statistic of 2.5549 with a p-value of 0.0797. Both of the test statistics are significant, which rejects the null of exogeneity and concludes that *REO* and *RWR* are endogenous regressors. Endogenous regressors will cause the OLS estimates to be inconsistent, thus the IV approach is an improvement over OLS.

Compared to the last estimation of the *REO* parameter 0.303 and the *RWR* parameter of 0.412, the new estimate shows that the inter-state economic migration are now more motivated by employment opportunities and less driven by relative compensation rate. This is consistent with what we observed of the U.S. economy since the recovery from recession began in mid-2009: while the economy is steadily healing from the recession and the unemployment rate is down, the weak labor market has put enormous pressure on wages and the real wages have barely grown. Economic migrants are now more responsive to shocks in employment opportunities relative to shocks in wage rate, mainly because wages are still stagnant.

2.5 Amenity Values

Amenity values of each state are estimated by the regional constant term λ^k. A positive regional constant indicates amenity value in the region is higher than the national average level; a negative regional constant indicates amenity value is lower than the national average. Based on the regional constants, REMI generates the compensating differentials to measures how much a worker need to be paid relatively to compensate the amenity loss by living in the region. The compensating differential is a relative to the nation measure with the national average equals to one. A compensation differential value higher than 1 indicates relatively poor amenities in the region, and vice versa.

$$Compensating\ Differential = e^{-1\times\frac{\lambda^k}{\beta}} \tag{11}$$

where

$\lambda^k = regional\ constant;$

$\beta = the\ average\ of\ model\ coefficients.$

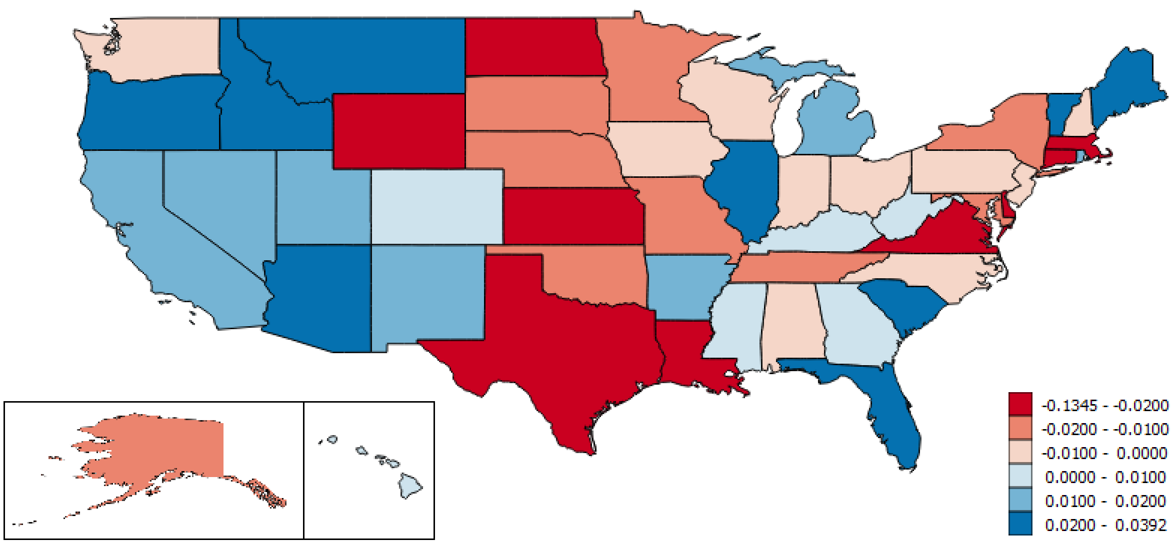
Table 2. Regional Constants and Compensation Differentials

State	Regional constant	compensating differential	State	Regional constant	compensating differential
Alabama	-0.0002	1.0005	Montana	0.0207	0.9488
Alaska	-0.0198	1.0516	Nebraska	-0.0142	1.0368
Arizona	0.0324	0.9211	Nevada	0.0103	0.9742
Arkansas	0.0137	0.9657	New Hampshire	-0.0017	1.0043
California	0.0190	0.9529	New Jersey	-0.0080	1.0206
Colorado	0.0030	0.9924	New Mexico	0.0148	0.9630
Connecticut	-0.0441	1.1185	New York	-0.0153	1.0395
Delaware	-0.0222	1.0581	North Carolina	-0.0005	1.0013
District of Columbia	-0.1345	1.4071	North Dakota	-0.0749	1.2095
Florida	0.0328	0.9201	Ohio	-0.0057	1.0145
Georgia	0.0075	0.9811	Oklahoma	-0.0114	1.0294
Hawaii	0.0063	0.9841	Oregon	0.0392	0.9054
Idaho	0.0360	0.9126	Pennsylvania	-0.0050	1.0127
Illinois	0.0257	0.9367	Rhode Island	0.0182	0.9548
Indiana	-0.0043	1.0110	South Carolina	0.0209	0.9483
Iowa	-0.0022	1.0055	South Dakota	-0.0132	1.0340
Kansas	-0.0289	1.0762	Tennessee	-0.0158	1.0409
Kentucky	0.0097	0.9757	Texas	-0.0249	1.0652
Louisiana	-0.0328	1.0868	Utah	0.0110	0.9725
Maine	0.0328	0.9201	Vermont	0.0310	0.9243
Maryland	-0.0158	1.0409	Virginia	-0.0415	1.1110
Massachusetts	-0.0463	1.1247	Washington	-0.0023	1.0060
Michigan	0.0167	0.9585	West Virginia	0.0032	0.9919
Minnesota	-0.0101	1.0260	Wisconsin	-0.0007	1.0019
Mississippi	0.0085	0.9787	Wyoming	-0.0323	1.0855
Missouri	-0.0106	1.0274			

Regional constants and compensating differentials are reported in Table 2. The state-level amenity values vary from -0.1345 to 0.0392. Compared to the regional constants that range from -0.1758 to 0.1348 from last estimation, the amenity values' effects on net economic migration are more similar across states and more close to the national average.

Figure 3 shows a map of amenity values measured by regional constants. The western and southeastern states have higher amenity values, while the eastern and mid-western states have relatively lower amenity values. Among the states, Oregon has the highest amenity value, while the District of Columbia has the lowest. A negative regional constant does not necessarily signal net outflow of economic migrants. Some states with negative regional constant (such as Louisiana, Alabama, and Virginia) has experienced net inflows of domestic migrants during 2009-2014. A negative regional constant only means that workers expect to be paid higher than average compensation to relocate to the area, given other things being equal. Similarly, a positive regional constant does not imply net inflows of migration. A sluggish local labor market may outweigh the positive amenity value and generate net outflows of population seeking employment.

Figure 3. Regional Constants from IV Estimates



Moreover, compared to last estimation, the majority of states remain in their “higher-than-average amenities” or “lower-than-average amenities” categories. Eight states that used to be higher amenity states has become lower amenity states; two states has switched from the lower amenity category to the other. One need to keep in mind that these amenity values or compensating differentials are relative measures of amenities in each state. Since the national average amenities change over time, changes in the relative state-level amenity values does not necessarily imply changes in absolute values of regional amenities in the same direction. Also, the amenity term in the structural equation only include those unobserved factors that have impacts on inter-state economic migration.

3. Conclusion

This study analyzes changes in the U.S. domestic migration patterns before and after the 2008 recession, and re-estimates the REMI migration equation using updated state-level data over the period 2009-2014. State-level net economic migration as share of the labor force is regressed on relative employment opportunity *REO*, relative real compensation rate *RWR* and regional constant of amenities λ^k. Omitted variables and the loop of causality between the dependent variable and the explanatory variables cause *REO* and *RWR* to be endogenous and biases the OLS estimator. To correct the endogeneity problem, the instrumental variable approach with fixed effects is utilized to estimate the migration equation. An industrial mixed employment growth rate variable *EMX* is generated to predict *REO* with federal military employment *FM*; and an industrial mixed average wage rate variable *WMX* is created to predict *RWR* with the fuel cost variable *fuel*.

The IV estimation results show that both the relative employment opportunity and the relative real compensation rate have significant positive effects on net inter-state economic migrants to the region. Relatively higher coefficient of *REO* indicates that inter-state migration are more motivated by employment opportunities rather than relative wage rate. This is in contrast with the estimation using 2001-2008 data, which suggests that the wage rate is a more important factor that affecting economic migrants' relocation decisions. In general, inter-state economic migrants are less responsive to shocks in the local job market after the recession.

The fixed effects in the structural equation measures the unobserved amenity value that influences the in or out of economic migrants of each state. The unobserved amenities could be any characteristics of the region that affect the net economic migration but are excluded from the structural equation. In addition to the economic factors, people who are making relocation decisions may also consider the climate, air quality, traffic condition, health services, recreation facilities, cultural diversity, etc. These factors are taken into account by the amenity term. The estimation results show that the amenity value through a regional constant that varies from -0.1345 to 0.0392 across states. Generally the western and southeastern states supplies amenities higher than the national average; the mid-western and northeastern states have relatively low amenity values. Whether a state gains or loses population seeking improvements in life standard is determined by combined effects of both the economic and the amenity factors.