College of the Holy Cross
CrossWorks

Economics Department Working Papers

Economics Department

12-1-2012

The Role of Expectations: An Application to Internal Migration

Robert Baumann
College of the Holy Cross, rbaumann@holycross.edu

Justin Svec
College of the Holy Cross, jsvec@holycross.edu

Francis Sanzari

The Role of Expectations: An Application to Internal Migration

By

Robert Baumann

Justin Svec

and

Francis Sanzari

December 2012

COLLEGE OF THE HOLY CROSS, DEPARTMENT OF ECONOMICS
FACULTY RESEARCH SERIES, PAPER NO. 12-05*

Department of Economics
College of the Holy Cross
Box 45A
Worcester, Massachusetts 01610
(508) 793-3362 (phone)
(508) 793-3708 (fax)

http://www.holycross.edu/departments/economics/website

*All papers in the Holy Cross Working Paper Series should be considered draft versions subject to future revision. Comments and suggestions are welcome.

The Role of Expectations: An Application to Internal Migration

Robert Baumann, Justin Svec, and Francis Sanzari*

December 14, 2012

Abstract

This paper examines the impact of unemployment on migration. In a theoretical model, we show that unemployment, per se, does not affect migration. Rather, migration only occurs when unemployment shocks force residents to update their expectations of the area's unemployment rate. Once these expectations change, migration reallocates labor to bring the economy back to equilibrium. To test this theory, we devise an empirical strategy using state level data in the U.S. from 2000 to 2010, we find strong empirical evidence that unemployment shocks outside of expectations have a far greater impact on migration than unemployment shocks that are within expectations.

JEL Codes: R23, J61, D8

Keywords: Migration, Unemployment, Expectations

*College of the Holy Cross, 1 College Street, Worcester, MA 01610.

1 Introduction

Labor market theory has long posited that economic factors induce migration across regions; see Lowry (1966) for an early exposition and Wozniak (2010) and Saks and Wozniak (2011) for an updated treatment. This migration helps reallocate resources to those regions where it is most valued and away from regions where it is least valued. One such factor considered to be an important driver of migration is unemployment: people, the theory holds, should immigrate to regions with low unemployment and emigrate from regions with high unemployment.

This prediction has been tested numerous times in the empirical literature. One common approach is to regress net migration on a region's unemployment rate and other covariates, using either state- or MSA-level data. For all its intuitive appeal, this prediction has not found consistent support in the data. As Greenwood (1975, 1997) notes, the coefficient on unemployment often has an unexpected sign or is not statistically significant.

There have been many suggestions why the empirical results do not confirm the theoretical intuition. One possible explanation is that, by using aggregate data, the studies were unable to capture the personal characteristics of interest; see, for example, Navratil and Doyle (1977). Another explanation is that simple panel regressions fail to capture the plethora of options available to a potential migrant, including the possible option of not migrating. This insight inspired a number of studies that employ a conditional logit model, such as Davies, Greenwood, and Li (2001) and Wozniak (2010).

In this paper, we present an alternative explanation. It is our contention that these unintuitive results stem from model misspecification. Specifically, we show that migration does not respond to the observed level of unemployment, per se, but rather people migrate due to changes in their *expectations* of unemployment across regions. The previous literature, in failing to account for these expectations, effectively averaged the effect of unemployment observations that did and did not change expectations.

To formalize this insight, we develop a parsimonious theoretical model of migration. This model assumes that a nation is composed of two regions, each of which is subject to an unemployment shock. The unemployment shock represents the probability that a resident in that region is unemployed. If the resident is unemployed, then she earns nothing; if she works, the resident earns her marginal product of labor. The residents of the nation are endowed with initial probability models that they believe characterize the regions' shock processes. In this environment, an equilibrium is defined as the distribution of population across regions such that no resident wants to migrate. This condition implies that the expected wages in the two regions must be equal to each other. As we will show, the initial population distribution depends on the residents' expectations of one region's unemployment rate relative to the other.

Given the initial equilibrium, we then consider two alternative scenarios. In the first scenario, we assume that each region is hit by an unemployment shock that falls within the residents' expectations. As we show, this event has no effect on migration because the residents' expectations do not change. In the second scenario, one region is hit with a level of unemployment that is outside of the residents' expectations. As this state of the world was not believed possible, the residents are forced to update their expectations about the shock process. This change in expectations then induces residents to migrate until the expected wages are once again equalized.

Using this theoretical model, we derive three testable predictions. First, the level of unemployment only affects migration if it alters expectations. If, instead, a shock occurs that does not change expectations, then migration should not occur. It is for this reason, we contend, that the previous literature, which typically ignores expectations, did not find a consistent impact of unemployment on migration. Second, if the shock improves (worsens) the residents' expectations about region i's unemployment shock, then residents immigrate into (emigrate from) region i. Third, unemployment shocks that cause greater changes in expectations lead to larger migration responses than shocks that cause smaller changes in expectations.

In our empirical section, we test these three predictions on U.S. state-level data from 2000 to 2010. To do this, we must differentiate between shocks that are within people's expectations and shocks that are outside. As these data are not readily available, we have created a simple empirical test that attempts to distinguish between these two types of observations. An unemployment shock is defined as within expectations if it is sufficiently close to past observations, where "close" is measured by k standard deviations from the mean of the past data. An unemployment shock is defined as outside expectations if it falls outside of that range. If k is large enough, then the shock can plausibly be considered unexpected, as that unemployment rate is far away from what has been observed historically.[1]

Given these definitions, we find strong evidence that unemployment observations outside of expectations have a much larger impact on migration than shocks within expectations. Further, we show that unemployment rates that are above expectations induce out-migration, while unemployment rates that are below expectations induce in-migration, just as the theory predicts. These results are robust to a number of alternative specifications. Our findings also suggest that unemployment shocks that are further away from people's expectations cause a greater impact on migration than shocks that are closer to (but still outside of) people's expectations. Finally, we find no evidence that a correlation between our expectations variable and large unemployment changes is creating a spurious result.

2 Theoretical model

A nation is divided into two regions, labeled region A and region B. The nation is populated by a continuum of identical residents. For simplicity, let the size of the continuum of residents be equal to unity. Call R^i the number of residents living in region $i \in \{A, B\}$, so that

$$R^A + R^B = 1 \tag{1}$$

These residents are perfectly mobile across region, and migration is costless. This latter assumption could be relaxed by assuming that migrants have to pay a fixed cost to move, but as this would not change our results, we do not add in this complexity.

Each region is subject to an exogenous unemployment shock, the only source of randomness in the model. This shock represents the probability that a resident living in region i is unemployed.[2] If a resident is unemployed, she earns nothing. This assumption is without loss of generality, as long as the amount an unemployed person earns is sufficiently small. If a resident is employed, she supplies labor inelastically and is paid a wage equal to her marginal product of labor. All employment contracts are assumed to be one period contracts. The production function in region i is

$$Y^i = \left(a - bL^i\right) L^i$$

where a and b are constants and L^i is the number of workers employed in region i.

The nation's residents are endowed with initial probability models that they believe characterize the regions' unemployment shock processes. The initial probability model for region i, labeled Π^i, is assumed to have two key characteristics. First, the distribution governing each region's unemployment shock is identically and independently distributed (i.i.d.).[3] Second, the probability model puts positive weight only on a finite number of unemployment values. Let n^i represent the number of possible states of the world that the residents believe are possible in region i.

The decision problem of each resident is to maximize her expected wage by deciding in which region to live. The expected wage of a resident in region i is

$$\begin{aligned} E\left[w^{i}\right] &= \sum_{s=1}^{n^{i}} \pi_{s}^{i}\left[u_{s}^{i}(0)+\left(1-u_{s}^{i}\right)\left(a-2bL_{s}^{i}\right)\right] \\ &= a\left(1-E\left[u^{i}\right]\right)-2bE\left[\left(1-u^{i}\right)L^{i}\right] \end{aligned}$$

where u_s^i is region i's unemployment rate if state s occurs. The residents will continue to migrate until both regions' expected wages are equal. When this condition is satisfied, no resident has the incentive to migrate. We define an equilibrium to be the distribution of residents across regions, $\left\{R^A, R^B\right\}$, such that

$$E\left[w^{A}\right]=E\left[w^{B}\right] \tag{2}$$

. To find the initial population distribution across regions, combine (2) with (1) and the restriction that, in any unemployment state s, $L_s^i=\left(1-u_s^i\right)R^i$. Doing this, the initial population distribution for region i is

$$R^{i*}=\frac{a\left(E\left[u^{j}\right]-E\left[u^{i}\right]\right)+2bE\left[\left(1-u^{j}\right)^{2}\right]}{2bE\left[\left(1-u^{i}\right)^{2}\right]+2bE\left[\left(1-u^{j}\right)^{2}\right]}$$

where $j \neq i$. Notice that the initial population in region i depends on the residents' expectations of each region's unemployment rate.

Given this initial equilibrium, we can now characterize the determinants of net migration, where net migration is defined as the percentage change in a region's population. To do so, we consider two alternative scenarios. In the first scenario, each region is hit by a shock that lies within the residents' expectations. In the second, one region is hit by a shock that is outside of the residents' expectations. For each scenario, we will calculate how migration responds to the observed level of unemployment.

Scenario 1:

Assume that both regions are hit by an unemployment shock that is within the residents' expectations. That is, the residents believed that the observed unemployment rates were possible, and so their occurrence gives the residents no new information. As a consequence, there is no need for the residents to update their expectations. Because the expectations are unchanged, the equilibrium population distribution does not change. This implies that there is no migration as a result of the unemployment shocks. This result highlights the model misspecification discussed above: the observed level of unemployment does not affect migration if the observed unemployment does not affect the residents' expectations.

Scenario 2:

Assume that region A is hit by an unemployment shock that is outside of the residents' expectations, while region B is hit by the same shock as in scenario 1. As region A's shock is of an unexpected magnitude, the nation's residents are forced to update their expectations. With a new set of expectations, the equilibrium population in region A is

$$R^{A\prime} = \frac{a\left(E\left[u^{B}\right] - \overset{\curlywedge}{E}\left[u^{A}\right]\right) + 2bE\left[\left(1-u^{B}\right)^{2}\right]}{2b\overset{\curlywedge}{E}\left[\left(1-u^{A}\right)^{2}\right] + 2bE\left[\left(1-u^{B}\right)^{2}\right]}$$

where $\overset{\curlywedge}{E}$ represents the updated expectation. Given the new population distribution we can calculate net migration as

$$Migration^{A} = \frac{R^{A\prime} - R^{A*}}{R^{A*}}$$

While the resulting value of net migration depends on how much the unemployment observation changes people's expectations, the important point to note is migration occurs in response to the change in people's expectations.

3 Empirical model

One central prediction emerges from the theoretical model presented above: only shocks to unemployment that are outside of expectations should impact migration across regions. In this section, we will test this prediction using US state-level data from 2000 - 2010.[4] We will also investigate two related hypotheses. We hypothesize that unemployment observations that are above (below) people's expectations cause people to emigrate from (immigrate into) the state. Also, we hypothesize that shocks that are further away from expectations, or unemployment observations that cause larger changes in expectations, should have a larger impact on migration than observations that are only slightly outside of people's expectations. As we will show below, we find substantial evidence for each of these predictions in the data.

Evaluating this theory requires an empirical measure of expectations. There are numerous possible strategies for obtaining this measure. First, we could assume that people, at some early date, were endowed with an initial expectation of each state's unemployment, $u^{i,e}$. This value could then be updated using some backward-looking process, yielding an expected value of the unemployment rate for each region and for each period. As an example, if we assume that people had adaptive expectations, they might update their expectations according to the following rule:

$$u_t^{i,e} = u_{t-1}^{i,e} + \lambda \left(u_t^i - u_{t-1}^{i,e} \right)$$

for some λ. A similar, and perhaps more common approach would be to endow people with a prior distribution over a wide support, and then allow Bayes' rule to update the prior. Second, we could look for direct measures of expectations. Analysts' forecasts of future unemployment rates, used for each state's budgeting process, are an example of this type of direct measure.

There are difficulties associated with each of these strategies. The first approach is problematic because, except under very specific and rare circumstances, all unemployment realizations would alter people's expectations. While we don't reject this as a possibility, this approach doesn't test our theory, as our model only updates expectations when the unemployment rate is unexpected. The second approach, obtaining a direct measure of expectations about future unemployment by month and by state, is problematic, as data like these are not easily available.

Given these challenges, we take a different approach, one that is consistent with the theoretical model presented above. We assume that, at some initial period t_0, people form state-specific unemployment expectations using the history of each state's unemployment observations. These expectations are ranges of unemployment rates that the residents believe to be possible in the following period. Specifically, each range is defined as the set of unemployment values that lie within $\mu_s \pm k\sigma_s$, where μ_s is the mean value of the unemployment observations in state s from $[t_0 - T, t_0)$, σ_s is the standard deviation of those same observations, and k is a constant. Note that in this definition T represents the number of periods of data people use to form their initial expectations, and k characterizes the width of the resulting expectations. Crucially, we assume that any unemployment observation that falls inside the range is within expectations, while observations that are outside of this range are outside expectations.

With these initial expectations defined, we now must discuss how these expectations evolve over time. To do this, consider an unemployment rate observed in state s at time t_0. This rate could be either inside or outside the range of initial expectations. If the unemployment rate is inside the range, then people expected that this unemployment rate was possible. Because of this, people do not have to update their expectations since the observed rate of unemployment matched their prior expectations. This means that people's expectations at date $t_0 + 1$ are the same as date t_0. The same logic applies at any date in the future: as long as the observed value of unemployment is within expectations, then people keep their previous period's expectations.

If the unemployment rate is outside the range, though, the observation was a surprise to the residents and forces them to update their expectations. We assume that at this point, people form their expectations in a similar manner as at the initial date t_0, except now they use the most recent T periods as their dataset. In this case, μ_s and σ_s would be calculated using $[t_0 + 1 - T, t_0]$ as data.[5] For use later, define $\max\_unemp_{st}(k)$ as the largest unemployment rate in the range of expectations for state s at time t using k as the multiple. Similarly, let $\min\_unemp_{st}(k)$ be the smallest unemployment rate in the range of expectations for state s at time t using k as the multiple.

There are a few points we would like to make about defining expectations in this manner. First, there might be many periods in a row in which expectations do not change. This would occur if the observed unemployment rates remain sufficiently close to the past observations. Second, while expectations might be static for long periods of time, they do change to reflect new information, as long as that information is sufficiently unlike what has been observed in the past. Table 1 illustrates this point using data from Colorado and $k = 2$. The rising unemployment rate in Colorado in late 2008 does not exceed expectations until it rises above 6.367%, which occurs in January 2009. For the next nine months, the unemployment rate remains outside of expectations. Meanwhile, expectations are adjusted to reflect the new information. By October 2009, expectations catch up to the new, higher unemployment rate in Colorado. In words, people adjust their expectations to incorporate a state's higher unemployment rate, but that this process takes time.

Third, if a state has witnessed a particularly volatile path of unemployment, expectations for that state are relatively broad. This, intuitively, seems reasonable because historical data are less informative about potential future unemployment rates. Fourth, for larger values of k, the range of expectations grows. This implies that it would take a more extreme observation of unemployment to induce migration. In our empirical analysis, we test a variety of values of k to check the robustness of our results to this multiple.

To initialize our algorithm and so determine the path of unemployment expectations over time and by state, we must make an assumption about T. This constant governs the number of previous periods that people use to form and update their expectations. A larger value of T would imply that people use a larger window of past data to form their expectations of future unemployment. In the results presented below, we assume that $T = 120$; that is, expectations are based on the past 120 months (or 10 years) of data. We have also tested the robustness of this value and have found that our results are robust to both a 5-year and a 20-year window.

With the state-specific paths of expectations in hand, we now form our two key independent variables. These variables indicate whether the observed rate of unemployment in state s at time t was within expectations and, if so, whether the observation was above or below the range. Let

$$flaghi\_k_{st} = \begin{cases} 1 \text{ if unemp}_{st} \geq \max\_unemp_{st}(k) \\ 0 \text{ o/w} \end{cases}$$

$$flaglo\_k_{st} = \begin{cases} 1 \text{ if unemp}_{st} \leq \min\_unemp_{st}(k) \\ 0 \text{ o/w} \end{cases}$$

With these definitions, if the observed unemployment rate was within expectations, then $flaghi\_k_{st} = flaglo\_k_{st} = 0$. However, if the unemployment rate was outside of the range of expectations, then either $flaghi\_k_{st}$ or $flaglo\_k_{st}$ should equal one, depending on whether the observation was above or below the range of expectations. To form the range and the two indicator variables, we have used unemployment data from Bureau of Labor Statistics, which compiles them using data from the Current Population Survey. This data tracks the unemployment rate by state and by month.

In addition to unemployment, we need data on net migration by state and by month. To obtain this, we use monthly population data for each state from January 2000 to September 2010. These data are created by the U.S. Census Bureau. With these values, we define migration as the percent change in population after mortality is removed. We use percent changes rather than the first differences since our geographic unit of observation (states) has a wide range of populations.

Upon examination of the data, we have decided to remove Louisiana and Mississippi from our dataset. These two states are outliers in both the migration and unemployment data due to Hurricane Katrina. Our results, though, are robust to this decision.

Table 2 presents the summary statistics of our data, including the sample means of $flaghi\_k_{st}$ and $flaglo\_k_{st}$, for several values of k. For each value of k, we have a larger number of high unemployment periods outside of expectations (i.e., $flaghi\_k_{st} = 1$) than low (i.e., $flaglo\_k_{st} = 1$). One explanation could be that our sample time frame includes the most recent recession in which unemployment rose quickly.

In our main empirical specification, we estimate the following fixed effects regression:

$$net\_migration_{st} = \beta_0 + \beta_1 * flaghi\_k_{s,t-1} + \beta_2 * flaglo\_k_{s,t-1} + \gamma_s + \delta_t + \epsilon_{st} \quad (3)$$

for several values of the multiple k. γ_s represents state-level fixed effects, and δ_t represents both yearly and monthly dummy variables. Including these will help capture any state- or time-invariant effects. Note that we use the lag of the unemployment variables to allow people time to migrate in response to changes in their unemployment expectations. Although we present our results using a one period lag, the results are similar for two-period and three-period lags. For comparison to the extant migration literature, we also run the same regression adding in the first difference of unemployment, lagged one period, as a covariate. All estimations use the heteroskedastic-consistent calculation of the standard errors. The estimates on the fixed effects are not presented below, but are available upon request, as are any of the other claims made above.

If our theoretical model finds support in the data, then we expect to find that the coefficients on the unemployment indicator variables are statistically significant. Further, we expect that the coefficient on $flaghi\_k_{s,t-1}$ ($flaglo\_k_{s,t-1}$) is negative (positive), as an unexpectedly large increase (decrease) in unemployment would deter (spur) immigration.

In the top half of Table 3, we present the numerical results from our main specification, assuming different values of the multiple, k. As can be seen, we find that $flaghi\_k_{s,t-1}$ has the theoretically correct sign and is statistically significant at the 1% level across all different values of k. Also, the coefficient on $flaglo\_k_{s,t-1}$ is generally positive and significant, consistent with the theory. The only time in which the coefficient on $flaglo\_k_{s,t-1}$ takes the opposite sign is when it is not statistically significant at $k = 2$. These results suggest that unexpectedly high observations of unemployment induce emigration from the state, while unexpectedly low observations of unemployment encourage immigration into the state.

Next, we modify the baseline specification to include $\Delta unemployment_{s,t-1}$. The results from these regressions are presented in the bottom half of Table 3. Similar to the above findings, $flaghi\_k_{s,t-1}$ is positive and statistically significant at the 1% level across all values of k. Further, whenever $flaglo\_k_{s,t-1}$ is statistically significant (which happens at $k = 1$ and $k = 3$), the coefficient is positive, as predicted. Finally, the coefficient on the change in unemployment is negative and statistically significant at the 10% level in each specification that includes this variable. This result runs counter to the simple theoretical model presented above, which predicts that the observed level (or change) in the unemployment rate is irrelevant to migration decisions, except insofar as it changes people's expectations. One reason for this divergence between the theory and the empirical results could be that we assumed that people expected an independent and identically distributed unemployment shock in the theoretical model. If, instead, people's expectations have some history dependence, we would find that the observed unemployment rate does influence migration. We would like to highlight, though, that the fit of the change in unemployment variable is generally worse than the $flaghi\_k_{s,t-1}$ and the $flaglo\_k_{s,t-1}$ variables and, as we will show, has a considerably smaller economic impact. In addition, including the lagged change in unemployment produces very small gains in overall explanatory power as measured by the r-squared.

To consider the economic impact of these estimates, consider a 0.1 percentage point increase in monthly unemployment. If the new unemployment value is within people's expectations, then (at $k = 1$) our results suggest that the median state (with a population around 3.12 million) would lose 22 people in the following month, an estimate that is statistically significant at ten percent. If, instead, the new unemployment value lies outside of people's expectations, then the median state would lose 586 people, or 0.019% of its population. While these numbers are small percentages of the state's population, note that they reflect a monthly impact on population. Also, the average length of time an unemployment rate is above expectations using $k = 1$ is 23.5 months. In other words, a median population state that spends the average length of time above its expected unemployment range loses an estimated 13,771 people or 0.44% of its population in just under two years.

Our estimates of the impact of an unexpectedly low unemployment shock on migration have a larger variance across k. The coefficient is positive and statistically significant for $k = 1$ and for $k = 3$, but not significant for $k = 2$. This may be driven by fewer observations of unexpectedly low unemployment periods. The former type of shock induces an in-migration that is about one-quarter as large as the latter type induces out-migration.

Exploring our numerical results further, we examine whether unemployment rates that are further away from the range of expectations have a greater impact on net migration than do unemployment rates that are close to, but still outside of, the range. Intuitively, this could occur because unemployment rates that are further away from the range of expectations cause larger changes in residents' expectations than unemployment rates closer to the range. To test this, we alter our definitions of the unemployment indicators to take into account the distance of the unemployment shock from people's expectations. Let

$$linearhi\_k_{st} = \begin{cases} \frac{unemp_{st}-\mu_s}{\sigma_s} \text{ if unemp}_{st} \geq \max\_unemp_{st}(k) \\ 0 \text{ o/w} \end{cases}$$

$$linearlo\_k_{st} = \begin{cases} \frac{unemp_{st}-\mu_s}{\sigma_s} \text{ if unemp}_{st} \leq \min\_unemp_{st}(k) \\ 0 \text{ o/w} \end{cases}$$

where μ_s, σ_s, $\max\_unemp_{st}(k)$, and $\min\_unemp_{st}(k)$ are defined as above. The variable $linearhi\_k_{st}$ captures not only whether the observed unemployment rate is above expectations, but also how far above those expectations, where the distance is measured in standard deviations from the mean. The opposite is true for $linearlo\_k_{st}$. With these new definitions, we regress net migration on the new unemployment indicators, each lagged one period, using year, state, and month dummies. For completeness, we also run the same regression adding in the change in unemployment, lagged one period.

As before, we predict that the coefficient on $linearhi\_k_{s,t-1}$ is negative and the coefficient on $linearlo\_k_{s,t-1}$ is positive. We also expect that an unemployment rate that is further away from expectations should have a greater impact on migration than an unemployment rate closer to expectations.

Table 4 presents the results from this new regression. The results are consistent with our previous findings and in line with our predictions. We find a negative and statistically significant coefficient on $linearhi\_k_{s,t-1}$ in every specification. Also, all specifications have a positive and statistically significant coefficient on $linearlo\_k_{s,t-1}$, except when $k = 2$. One difference from the previous results is that the change in unemployment variable is no longer statistically significant at $k = 1$.

Unsurprisingly, our point estimates are closer to zero than before. This reduction was expected because we no longer average the impacts of all unexpected unemployment rates, but rather separate them based on their distance from the mean. These point estimates suggest that, indeed, more extreme unexpected unemployment shocks have greater effects on migration than less extreme unexpected unemployment shocks. To see the size of this difference, consider a median population state whose unemployment rate has jumped outside of expectations. If the unemployment rate is one standard deviation above the mean of the range of expectations (using $k = 1$), then 273 people will migrate away from the state in the following month. But, if the state's unemployment rate is two standard deviations above, the state's out-migration will double.

Once again, the migration response for low, unexpected unemployment rates is more volatile relative to high, unexpected unemployment rates. The amount of this difference depends on the level of k. For $k = 1$, unexpectedly low unemployment rates have about half the migration response compared to unexpectedly high unemployment rates. For $k = 3$, the migration response is larger for low, unexpected unemployment rates in absolute terms. More study is needed to examine this point, however, since we have few observations of large, unexpected improvements in a state's unemployment rate in our time frame.

We take the above results to be strong evidence that the role of expectations are central to understanding the impact of unemployment on net migration. In particular, unemployment observations that are outside of expectations have a much greater impact on migration (by a factor of approximately 25) than do observations within expectations. Moreover, the further the observation is from the range of expectations, the greater the impact on migration.

There is, though, a potential concern about the interpretation of our results. Perhaps we are not capturing how changes in expectations affect net migration in our analysis, but rather we are capturing the fact that large changes in unemployment drive migration. Under this alternative interpretation, our expectations variable merely proxies for large unemployment changes. To examine whether this is the case, we construct a new set of unemployment indicator variables. Let

$$\begin{aligned} hi\_k_{st} &= \begin{cases} 1 \text{ if } \Delta unemp_{st} \geq 0.2 \text{ and } flaghi\_k_{st} = 0 \\ 0 \text{ o/w} \end{cases} \\ lo\_k_{st} &= \begin{cases} 1 \text{ if } \Delta unemp_{st} \leq -0.2 \text{ and } flaglo\_k_{st} = 0 \\ 0 \text{ o/w} \end{cases} \end{aligned}$$

These variables isolate periods with large unemployment changes that do not trigger the flag variables. That is, these new variables collect the largest changes in unemployment (the unemployment change must be greater than 0.2 in absolute value) in which the resulting unemployment rates are still within people's expectations.

Table 5 presents summary statistics for hi_k_{st} and lo_k_{st}. Of the 446 observations where $\Delta unemp_{st} \geq 0.2$, about 80% overlap with periods in which unemployment is above the expected range, using $k = 1$. The amount of overlap drops as k increases because there are fewer periods outside of expectations by definition. Finally, there is no overlap between periods where unemployment fell more than 0.2 percentage points and the unemployment rate was below expectations. This second fact immediately offers suggestive evidence supporting our interpretation that changes in expectations drive net migration, not large changes in unemployment.

With these variables, we regress net migration on $hi\_k_{s,t-1}$, $lo\_k_{s,t-1}$, and our state, year, and month dummies. We also could include the $flag$ or $linear$ variables, but we omit these results since they do not substantially change our findings. If our theoretical model is correct and expectations are central to migration, we would expect that the coefficients on $hi\_k_{s,t-1}$ and $lo\_k_{s,t-1}$, to be statistically indistinguishable from zero. Instead, if the alternative view is correct and migration depends only on large changes in unemployment, then we expect that $hi\_k_{s,t-1}$ and $lo\_k_{s,t-1}$ will be statistically significant and have the right signs.

Table 6 presents our results from this regression. As can be seen, the coefficients on $hi\_k_{s,t-1}$ and $lo\_k_{s,t-1}$ are statistically indistinguishable from zero in every specification. We believe that these results, combined with Tables 3 and 4, provide substantial support for our interpretation that unemployment expectations are central to understanding migration, while providing no evidence that large changes in monthly unemployment drive migration. It is this latter point that strengthens our earlier assertion that the previous literature's regression of net migration on the observed values of unemployment led to model misspecification and hence, mixed results.

4 Conclusion

In this paper we develop a parsimonious theoretical model that shows that migration only occurs when an unemployment rate causes expectations to change rather than occurring whenever the unemployment rate changes. That is, only in periods where the observed unemployment rate is outside of expectations is there a familiar linear effect on migration. Our empirical model verifies this prediction. A change in the monthly unemployment rate produces vastly different effects on migration depending on whether the new unemployment rate is within expectations or not. Using $k = 1$ and a median population state, a 0.1 percentage point increase in the monthly unemployment rate corresponds to an out-migration of 22 people compared to 586 people if the shock pushes the new unemployment rate above expectations. We also find in-migration for periods where the unemployment rate is unexpectedly low, though this estimate fluctuates depending on the breadth of expectations. Finally, our robustness check verifies that our conclusions are not spurious due to the correlation between our definition of expectations and large changes in unemployment.

Our findings provide a potential explanation for regions of persistent poverty, e.g. Appalachia or some inner-cities, where out-migration occurs slower than expected. For residents of these areas, a high unemployment rate is likely within their expectations since it has occurred often in the past. As a consequence, residents of Appalachia or these inner-cities don't migrate out because the high unemployment rate didn't change their expectations.

We also believe that our approach to modeling and empirically measuring expectations could have a wide range of applications outside of the migration literature. For example, consider the random variable associated with the growth rate in the average housing price in the U.S. Because that random variable was consistently positive prior to 2008, many investors and regulators behaved as if it could not become negative. This state of the world was outside of their expectations. Only after the random variable was observed to be negative and investors were forced to update their expectations did their behavior change.

There remains some ambiguity about our model of expectations. Specifically, how do consumers transform information into expectations for the future? For example, how far back in the past do they use information (that is, what is the true value of T)? How wide are their expectations (k)? Do they update expectations each period or only when shocks are outside of expectations? Fortunately, our estimates are largely robust across these questions, but we believe that understanding the formation of expectations is the next step in specifying the relationship between unemployment and migration.

References

[1] Davies. P, M. Greenwood, and H. Li (2001). A Conditional Logit Approach to U.S. State-to-State Migration. *Journal of Regional Science* 41(2), 337-360.

[2] Greenwood, M. (1975). Research on Internal Migration in the United States: A Survey. *Journal of Economic Literature* 13(2), 397-433.

[3] Greenwood, M. (1997). Internal Migration in Developed Countries. *Handbook of Population and Family Economics*. Amsterdam: Elsevier, 647-720.

[4] Lowry, I. S. (1966). *Migration and Metropolitan Growth*. San Francisco: Chandler.

[5] Navratil, F. and J. Doyle (1977). The Socioeconomic Determinants of Migration and the Level of Aggregation. *Southern Economic Journal* 43, 1547-1559.

[6] Saks, R and A. Wozniak (2011). Labor Reallocation over the Business Cycle: New Evidence from Internal Migration. *Journal of Labor Economics* 29(4), 697-739.

[7] Wozniak, A. (2010). Are College Graduates More Responsive to Distant Labor Market Opportunities? *Journal of Human Resources* 45(4), 944-970.

Notes

[1] We do not claim that this is the only definition of shocks that are outside or within expectations. Instead, our assumed definition is merely one of many possible definitions. As such, our empirical results should be viewed as supportive but not conclusive evidence on the implications of our theoretical model.

[2] Although this model is written without explicit reference to time, we implicitly assume that all workers living in a region have an equal probability of being employed in the following period, where that probability is equal to one minus the unemployment rate.

[3] Our assumption that residents believe that the regions' unemployment rate draws are i.i.d. is not innocuous. We could have, instead, assumed that there was some history dependence in people's expectations (the shock could be characterized by a random walk process, for example). This alternative assumption would imply that all unemployment observations alter people's expectations. While we do not reject this as a possibility, the goal of our theoretical model is to show that only unemployment observations that change people's expectations cause migration. As such, we needed a process in which some observations change people's expectations while others do not.

[4] To be more precise, we have unemployment data by month and by state from 1990-2010 and net migration data by month and by state from 2000-2010. We use the pre-2000 unemployment data to create our measure of the initial expectations of the state's unemployment rates for the year 2000. Then, with the algorithm detailed below, we use the unemployment data to update expectations and to determine which unemployment observations are within those expectations and which are outside. With this information, we test that prediction of our theoretical model that only shocks that change expectations affect net migration.

[5] As a robustness check, we considered an alternative approach to forming people's unemployment expectations. In that approach, we assumed that people updated their expectations each period using a rolling window of past unemployment observations, regardless of whether the unemployment observation was outside or inside expectations. This alternative approach yielded qualitatively similar results to the model we currently present. The results from that alternative approach are available on request from the authors.

Table 1: Expectations Definition Example: Colorado

year	μ_{st}	σ_{st}	***lower bound***	***upper bound***	***unemp. rate***	***outside of exp.?***
Sep 2008	4.431	0.968	2.495	6.367	5.1	no
Oct 2008	4.431	0.968	2.495	6.367	5.4	no
Nov 2008	4.431	0.968	2.495	6.367	5.7	no
Dec 2008	4.431	0.968	2.495	6.367	6.2	no
Jan 2009	4.539	1.139	2.261	6.816	6.7	yes
Feb 2009	4.571	1.146	2.279	6.862	7.3	yes
Mar 2009	4.606	1.162	2.282	6.932	7.7	yes
Apr 2009	4.648	1.186	2.276	7.019	8.1	yes
May2009	4.690	1.218	2.254	7.127	8.2	yes
Jun 2009	4.732	1.252	2.228	7.237	8.3	yes
Jul 2009	4.774	1.287	2.200	7.349	8.1	yes
Aug 2009	4.814	1.316	2.182	7.446	7.9	yes
Sep 2009	4.853	1.338	2.177	7.529	7.7	yes
Oct 2009	4.853	1.338	2.177	7.529	7.5	no
Nov 2009	4.853	1.338	2.177	7.529	7.4	no
Dec 2009	4.853	1.338	2.177	7.529	7.3	no

Note: This example uses k = 2. The formula for the range of expectations is $\mu_{st} \pm 2\sigma_{st}$. To determine whether an unemployment rate is outside the range of expectations, it is compared against the previous month's upper and lower bounds.

Table 2: Summary Statistics

	mean (standard deviation)	
population growth rate	0.092 (0.006)	
Δ unemployment rate	0.036 (0.139)	
k = 1	***frequency***	***percent***
outside high	1,676	26.51%
outside low	1,338	21.17%
***k* = 2**		
outside high	840	13.29%
outside low	202	3.20%
***k* = 3**		
outside high	354	5.60%
outside low	34	0.54%

Note: There are 6,321 state-month observations.

Table 3: Estimation – Intercept Effect
(p-values in parentheses)

	***k* = 1**	***k* = 2**	***k* = 3**
estimation without Δ unemployment			
$flaghi\_k_{s,t-1}$	-0.01854 (<0.001)	-0.01709 (<0.001)	-0.01817 (<0.001)
$flaglo\_k_{s,t-1}$	0.00632 (0.001)	-0.0029 (0.491)	0.0208 (0.043)
r-squared	0.6848	0.6815	0.6815
estimation with Δ unemployment			
Δ unemployment $_{t-1}$	-0.00702 (0.088)	-0.01331 (0.001)	-0.01044 (0.009)
$flaghi\_k_{s,t-1}$	-0.01816 (<0.001)	-0.01683 (<0.001)	-0.01748 (<0.001)
$flaglo\_k_{s,t-1}$	0.00597 (0.003)	-0.00356 (0.395)	0.02021 (0.050)
r-squared	0.6850	0.6821	0.6819

Note: These estimations also include year dummies, month dummies, and state-level fixed effects.

Table 4: Estimation – Linear Effect
(p-values in parentheses)

	***k* = 1**	***k* = 2**	***k* = 3**
estimation without Δ unemployment			
$linearhi\_k_{s,t-1}$	-0.00875 (<0.001)	-0.00564 (<0.001)	-0.00526 (<0.001)
$linearlo\_k_{s,t-1}$	0.00452 (0.001)	-0.00040 (0.829)	0.00829 (0.016)
r-squared	0.6869	0.6818	0.6819
estimation with Δ unemployment			
Δ $\text{unemployment}_{t-1}$	-0.00157 (0.699)	-0.01141 (0.004)	-0.01001 (0.012)
$linearhi\_k_{s,t-1}$	-0.00869 (<0.001)	-0.00548 (<0.001)	-0.0051 (<0.001)
$linearlo\_k_{s,t-1}$	0.00446 (0.001)	-0.00064 (0.732)	0.00812 (0.018)
r-squared	0.6869	0.6822	0.6823

Note: These estimations also include year dummies, month dummies, and state-level fixed effects.

Table 5: Robustness Check – Summary Statistics

	frequency	
Δ unemployment > 0.2	446	
Δ unemployment < -0.2	41	
Overlap	***unemployment above expectations***	***unemployment below expectations***
k = 1	357 (80.04%)	0
k = 2	191 (42.83%)	0
k = 3	114 (25.26%)	0

Note: Overlap refers to the number of observations where *Δ unemployment* > 0.2 and the unemployment rate is outside of expectations for a given value of *k*. The percentages refer to the percent overlap and are taken out of 446.

Table 6: Robustness Check – Estimation
(p-values in parentheses)

	***k* = 1**	***k* = 2**	***k* = 3**
estimation without Δ unemployment			
$hi\_k_{s,t-1}$	-0.00068 (0.861)	-0.00346 (0.158)	-0.00339 (0.128)
$lo\_k_{s,t-1}$	0.00049 (0.933)	0.00048 (0.935)	0.00043 (0.942)
r-squared	0.6800	0.6801	0.6801
estimation with Δ unemployment			
Δ unemployment	-0.01610 (<0.001)	-0.01550 (0.001)	-0.01602 (0.001)
$hi\_k_{s,t-1}$	0.00268 (0.501)	0.00006 (0.982)	0.00064 (0.779)
$lo\_k_{s,t-1}$	-0.00590 (0.333)	-0.00568 (0.353)	-0.00587 (0.339)
r-squared	0.6807	0.6807	0.6807

Note: These estimations also include year dummies, month dummies, and state-level fixed effects.