“Environmental migration beyond the Dust Bowl in the 1930s”

Myron P. Gutmann, Daniel Brown, Angela Cunningham, Susan Hautaniemi Leonard, Jani S. Little, Jeremy Mikecz, Paul Rhode, Seth Spielman, Kenneth M. Sylvester.

Abstract

This paper analyzes in detail the role of environmental and economic shocks in the migration of the 1930s. The 1940 U.S. Census of Population asked every inhabitant where they lived five years earlier, a unique source for understanding migration flows and networks. Earlier research documented migrant origins and destinations, but we will show how short term and annual weather conditions at sending and receiving locations in the 1930s explain those flows, and how they operated through agricultural success. Beyond demographic data, we use data about temperature and precipitation, plus data about agricultural production from the agricultural census. The widely known migration literature for the 1930s describes an era of relatively low migration, with much of the migration that did occur outward from the Dust Bowl region and the cotton South. Our work about the complete U.S. will provide a fuller examination of migration in this socially and economically important era.

In this paper we attempt for the first time to understand in detail the role of environmental and economic shocks in the migration of the 1930s, at the individual level, and for the United States as a whole. The 1940 Census of Population asked every inhabitant of the U.S. where they had lived five years earlier, providing a unique source for understanding migration flows and networks. Earlier research documented migrant origins and destinations, but this paper will go further, showing whether short term and annual weather conditions at sending and receiving locations in the 1930s explain those flows and networks, and the extent to which they operated through agricultural success. In addition to the demographic data, in this paper we make use of data about temperature and precipitation, as well as tabulated data about agricultural operations and production from the U.S. agricultural census. The key questions emerge from the widely known migration literature for the 1930s, which tells us that this was an era of relatively low migration (compared with the 1920s and the 1940s, for example), with much of the migration that did occur outward from the Dust Bowl region and the cotton South. The conclusions to the paper should provide a good starting point for a fuller examination of migration in this socially and economically important era.

People on the move are an enduring image of the U.S. in the 1930s, from photographs (Agee & Evans, 1941), literature (Steinbeck, 1939), and history (Egan, 2006; Gregory, 1989). People did move in the 1930s, spurred by the economic difficulties of the Great Depression, by heat and drought in the middle of the decade, and by a multitude of other pressures. Despite the lore of the Dust Bowl and the “Great Migration” from the South to the North, the volume of internal U.S. state-to-state migration was not large from the early 20th century until after World War II (around 12% made inter-county moves, (U.S. Bureau of the Census, 1946)). Moderate interstate migration rates belie a continuing mobility made up of streams of people moving relatively short distances from one type of community to another (Bogue, Shryock, & Hoermann, 1957; Ferrie, 2006; Hall & Ruggles, 2004; U.S. Bureau of the Census, 1946). Nonetheless, migration was a frequent subject of discussion, important enough that the U.S. Census tracked migration for the first time in 1940, asking exactly where every person enumerated had lived five years earlier (U.S. Bureau of the Census, 2002), providing a reliable way to measure a critical element of demographic change at an important moment in time. With this wealth of data, research about
migration in the 1930s has explored many questions but left even more unanswered (Bogue & Hagood, 1953; Bogue et al., 1957; Boustan, Fishback, & Kantor, 2010; Boustan, Kahn, & Rhode, 2012; Fishback, Horrace, & Kantor, 2006; Hornbeck, 2012; Lively & Taeuber, 1939; Long & Siu, 2013; McLeman, 2013; McLeman & Smit, 2006; Tolnay, White, Crowder, & Adelman, 2005; U.S. Bureau of the Census, 1946; White, 2005; White, Crowder, Tolnay, & Adelman, 2005).

Migration has been understood through a long theoretical literature (Greenwood, 1997; Lee, 1966; Massey, 1990; Massey et al., 1993, 1998; Ravenstein, 1885, 1889; Roy, 1951). Much of this literature is about international migration, but its core elements also apply to internal migration. Neoclassical economics provides the dominant perspective, informed by the contextually sensitive additions of the New Economics of Migration, and enriched by the explicitly multilevel models implied by Massey’s notion of Cumulative Causation and by the descriptive richness of Migration Systems Theory (Bakewell, 2013; Fussell, Curtis, & DeWaard, 2014; Massey, 1990; Massey et al., 1993). In these approaches, a person’s likelihood of migrating is a function of his or her individual characteristics plus the context at both origin and destination, including distributions of income, land, and human capital; the organization of agriculture and industry; public policy; and cultural elements, such as ethnic and religious structure (Fishback et al., 2006; Massey et al., 1993). Migration also takes place in a universe of social networks. These processes are spatial, in that the attributes of people in their spatial context are relevant, so we aim to see migration as something that takes place across distances and through pathways that can be connected, and not as an aspatial analysis of people in unconnected and location-free economic, social, or political contexts.

Most research sees migration as a tension between pushes and pulls. Disaster-related demographic theory builds on migration systems analysis to identify the migratory streams that existed prior to the disaster and to gauge whether a given shock transforms the pre-existing migration system (Black et al., 2011; Fussell et al., 2014; McLeman, 2013). It also adds the concept of vulnerability to the determinants of migration (Adger, 2006; McLeman, 2013; McLeman & Smit, 2006). The vulnerability paradigm focuses on the exposure of people to stress, based on the economic and ecological systems they inhabit. Integrated assessments of vulnerability have enabled research in both short-term moves and long-term reorganizations of human-environment systems (Adger, 2006; Berkes, Carl, & Johan, 1998; Black, Arnell, Adger, Thomas, & Geddes, 2013), asking whether a specific “demographic signature of disaster” exists (DeWaard, Curtis, & Fussell, 2014). In the U.S., the Dust Bowl story has driven research about environmental migration in the 1930s, often without enough attention to environmental stress in other areas. Drought and land degradation were not limited to the southern plains. Extreme heat led to agricultural stress elsewhere (Giesen, 2011; Gregory, 2005; McEwan et al., 2014; Olmstead & Rhode, 2008), as demonstrated in Figures 1-4, which show changes in agricultural conditions across the U.S. between 1930 and 1935.

Our analysis goes beyond environmentally-driven agricultural shocks, however. The U.S. was in the midst of depression in the 1930s, with low wages, weak economic performance, and high unemployment. We also ask whether people left places with comparatively poor economic conditions (being pushed) to go to places with better conditions (being pulled), and to do so in a way that reveals the characteristics of places as drivers of migration.

The migration literature is heavily focused on the idea first raised by Roy (1951) that migrants self-select for upward mobility. Borjas (1987) expanded selection theory by arguing that it is the more highly skilled who migrate. Kanbur and Rapoport (2005) argue that selectivity by education is key. The important idea to understand is that individual attributes are as important as those of context in determining who migrates, from which communities, and where they go. Our research doesn’t include information about each person’s or family’s individual vulnerability.
in 1935, but our knowledge of some of their characteristics, including age, sex, educational attainment, and family and marital status permits the analysis in this paper.

**Data and Methods**

The analyses for this paper call for a variety of data, at different spatial and temporal scales. The 1940 U.S. Census of Population full count data have a wealth of information about individuals, including demographic, social and cultural, economic, and location data (Ruggles et al., 2010), and constitute the primary data source for our analysis. At this time, the full-count 1940 data are available in two forms, and we make use of both of them. The University of Minnesota has released a preliminary version of the 1940 data in IPUMS coded format, which includes every person in the U.S., but does not include every variable from the census. For example, at the moment it does not include the city and county of residence in 1935. In addition, they have also made a restricted release of a version of the 1940 data that includes the detailed response text for every person. We have merged these data sets, and coded -- to the extent possible -- city and county from the 1935 textual residence variable. These data provide us with dependent variables as well as important individual-level independent variables including age, educational attainment, and marital status (although we cannot be certain that education and marital status in 1940 are perfectly informative of the individual’s status in 1935).

We also make use of county-level data from various U.S. population, economic and agricultural censuses. On the economic side we make use of well-known data from economic censuses about employment and industry, home ownership, and farm dependence for various years prior to 1940 (Haines, 2010; Haines, Fishback, & Rhode, 2014; U.S. Bureau of the Census, 1937; U.S. Dept of Labor, 2009). For agriculture, we have data about the level of crop production and livestock inventories from the 1930, 1935, and 1940 agricultural censuses. Among other analytic tools, we have developed a measure of agricultural shocks that will be useful for the diversity of U.S. agriculture in the 1930s by identifying the three most commonly planted crops in each county and measuring the deviation of production from long-term trends, combined with reports of failed and idle acres (Figure 5). We also derive environmental data affecting agriculture (precipitation, temperature, soil quality) and about environmental shocks from a number of publicly available data collections, including the PRISM high resolution climate data (Daly, Gibson, Taylor, Johnson, & Pasteris, 2002; Daly et al., 2008), a self-calibrating Palmer Drought Severity Index (PDSI) gridded data set for the 20th century U.S. (Mitchell & Jones, 2005; National Center for Atmospheric Research Staff eds., 2014), and land capability class data from the Soil Survey Geographic (SSURGO) database of the Natural Resources Conservation Service (NRCS) (Sylvester, Brown, Deane, & Kornak, 2013; Sylvester & Rupley, 2012).

Our analysis takes advantage of recent developments in multilevel models, like Alison’s (2009) hybrid fixed-effects model, which combines the accuracy of fixed-effects and the efficiency of random-effects estimators. A conventional fixed-effects model assumes that all unobserved covariates may be correlated with observed variables, and controls for unobserved heterogeneity by controlling for each unit of analysis. In an analysis of county-to-county migration in the U.S. (the relatively simple approach taken in specific aim 2), a fixed effects model would require the introduction of roughly 3100 dummy variables. The hybrid model duplicates the unbiased estimation of the fixed-effects model by either using mean deviation predictors (mean-centering) in time-varying designs, or relying only on stable (non-time-varying) predictors in multi-level cross-sectional or conditional logistic designs, which are appropriate for phenomena like migration. An example of a transition modeled might be a non-repeated event. The ‘case-time-control’ design (Suissa, 1995) offers a way to model such a working dependent variable as a discrete function of time during a control period (1935 to 1940), in which time dependence co-varies with stable predictors like completion of formal education, the birth of a
The working model is conditional logistic regression estimation

\[
\log \left( \frac{P_{it}}{1 - P_{it}} \right) = \alpha_i + \beta_1 H_{it} + \beta_2 t + \beta_3 t^2
\]

where \( H_{it} \) is a dummy variable for the event of interest and \( P_{it} \) is the probability of the covariate event occurring within a specified number of years prior to the migration event. The odds ratios are interpreted as the effect of the other life event on the migration. There is some advantage to limiting the estimation to those who experience the event (Allison, 2009), but we will also explore designs which include all censored cases in the population (those persons who did not experience the transition).

References.


Map 5. Weighted percent change in crop production for the top three crops in each county, 1930-1935.