The Dynamics of Mass Migration
Estimating the Effect of Income Differences on Migration
in a Dynamic Model of Discrete Choice with Diffusion*

Yannay Spitzer†

Population Studies and Training Center, Brown University

October 31, 2014

Abstract

During the period 1881-1914, approximately 1.5 million Jews immigrated to the U.S. from the Pale of Settlement in the Russian Empire. The data generated by this event can help explain the puzzling pattern of transatlantic mass migration: while time-series evidence shows that levels of migration were very volatile and highly sensitive to business-cycle fluctuations, there is little cross-sectional evidence for a systematic effect of income on migration—poorer countries did not always send more emigrants than wealthier countries. I explain this puzzle by using a newly constructed and unique data set, linking Ellis Island individual arrival records of hundreds of thousands of Russian Jews to information from the 1897 Russian census on their towns and districts of origin. I document the evolution of their migration networks using data on the incorporation of hometown-based associations, and capture local push shocks from comprehensive data on crop yields in Russia. Using a dynamic model of discrete choice with unobserved heterogeneity and an underlying networks diffusion process, I estimate the short-run effect of income shocks on migration, and show how the long-term effect of income levels could in principle be identified, net of the effects of network and of short-term income fluctuations. I find that the strong reaction of migration flows to business cycles can be largely attributed to individuals optimally timing their migration—temporary shocks to migration were offset in the long run by delayed migration. Finally, I provide evidence affirming the disputed diffusionist view, that the entry of the European periphery to mass migration was delayed for decades due to the time required for migration networks to diffuse across the continent.

*I thank my committee members Joel Mokyr (chair), Igal Hendel, and Joseph Ferrie for their endless support and advice. I am grateful for comments and suggestions from members of the economic history community of Northwestern University, participants in the Northwestern economic history seminars, members of the department of economics at the Hebrew University of Jerusalem, to Gautam Gowrisankaran and Ariell Zimran. This work has benefited from material support from the Economic History Association’s Sokoloff Fellowship, the Crown Family Fellowship of Northwestern University’s Crow Center for Jewish Studies, Northwestern University Center for Economic History, Northwestern University Graduate School’s Graduate Research Grant and Technology-Enhanced Doctoral Research Award, as well as by several periods of hosting at the department of economics at the Hebrew University of Jerusalem. Data used in this paper was constructed with Gennady Polonetsky and Marcy Blattner. Research assistance was provided by Amanda Lerner, Annie Weiss, Elijah Wilson, and Victoria Zak. I am thankful to Peg Zitko and the Statue of Liberty—Ellis Island Foundation for providing the Ellis Island arrival records data, and to Andre Zerger and Joel Perlmann for sharing GIS data on the Russian Empire.

†Email: yannay.spitzer@brown.edu, personal website: yannayspitzer.net.
1 Introduction

Between 1881 and 1914, 1.5 million Jewish immigrants arrived to the United States from the Pale of Settlement, the 25 western provinces of the Russian Empire in which Jewish population was generally allowed to reside and that was home to the world’s largest Jewish population (about 5 millions, as of 1897). Along with the Russian Revolution, the Holocaust, and the formation of the State of Israel, this exodus was one of a few colossal events that revolutionized world Jewry in the modern era. It was also an exemplary case of the way mass immigration has built up America, adding up a stock of initially empty-handed immigrants who soon rose to contribute their share in making the United States a leader in entrepreneurship, science and arts.

The purpose of this paper is to use the case of the Jewish migration from Russia and the uniquely rich data it produced to explain the puzzling pattern of pre-WWI European transatlantic mass migration: As can be seen in Figure 1a, the concurrence of downturns in immigration and U.S. recessions was a repeating phenomenon. The major immigration peaks corresponded to the onset of the big recessions of 1873, 1884, 1893, and 1907, with troughs leveled well below half their preceding peaks. As shown in Figure 1b, the yearly fluctuations in immigration were very large. In the period 1851-1914 the median absolute yearly logarithmic change in U.S. immigration was 0.21. However, poorer countries did not always produce more emigration. Figure 2 demonstrates how a negative cross-section correlation existed between real wages and emigration from Europe in the period 1890–1913, but not during the previous two decades.1

It is puzzling that the relatively small changes in expected lifetime income brought about by business cycle fluctuations produced such huge swings in migration, while much larger differences across countries failed to consistently correlate with differences in the national rates of emigration. In other words, if the income elasticity of migration appears so large in the time-series, why did the mass migration from the poorer European periphery, including Italy, The Habsburg Empire, and the Russian Empire, lagged behind and took several decades to catch up and surpass the migration from the wealthier western European countries?

Using the case of the Jewish migration from Russia, this paper revises the debate on the European pattern of transatlantic mass migration, and advances a dual explanation: (a) The sharp procyclical- cality of migration, with respect to destination business cycles, was driven by the countercyclicality of the costs of migration, and it reflects a sensitivity of the timing of migration rather than an actual strong long-run income elasticity of migration. (b) The late arrival of transatlantic migration from the European periphery was a result of a gradual process of spatial diffusion of migration networks across the European continent; the demand for migration in southern and eastern Europe may have been very large decades earlier, but it was inhibited due to the lack of personal links to friends and relatives who had already migrated.

1 For a comprehensive discussion of pre-WWI patterns of transatlantic mass migration, see Hatton and Williamson (1998).
Neither parts of the explanation are new. The former is rather consensual. The issue was pointed out by Gould (1979), and formalized in a time-series estimation model by Hatton (1995) and Hatton and Williamson (1998). Boustan (2007) and Bohlin and Eurenius (2010) applied this model to the cases of the Jewish migration from Russia and the Swedish emigration. This paper contributes by proposing a dynamic model of migration based on micro-foundations, and by using previously unavailable fine resolution data; these help to overcome problems of identification that prevailed in previous attempts to separately identify the long-run effect of income and the short-run effect of business cycle shocks.

The latter part of the explanation, however, is a minority view. It was advocated by Gould (1980b) and Baines (1995), but lack of sufficient data and estimation methods has left the verdict on this hypothesis still pending. Following Spitzer (2014), that demonstrated that spatial diffusion must have played a crucial role in the evolution of the Jewish migration from Russia, the estimation model used in this paper explicitly incorporates an underlying process of networks diffusion that interacts with the demand model of migration. In addition to formally testing the diffusionist theory and quantifying the role of networks, modeling the diffusion process is crucial for correctly identifying the parameters of the demand model, and it also helps in making it computationally feasible.

Records of individual arrivals of hundreds of thousands of Russian Jews to Ellis Island during the fiscal years (FY) 1900–1914 were linked to their towns of origin in the Pale of Settlement, and matched to data on the Jewish population in the Russian Empire from the 1897 Census. I used these to create a high-resolution panel of cohort-district-year migration counts from more than two hundred Russian districts. To account for the long-run evolution of migration networks, I linked records of incorporation of 1,476 landsmenschaftn—Jewish hometown-based associations in New York City—to their respective towns of origin in Russia. These were aggregated to form a district-year panel over the period 1861–1920, standing as rough measures proxying for the local exposure to migration. I complemented standard time-series of American real wages and Russian income per capita with a new province-year level panel of exogenous shocks to agricultural output in the Russian Empire over the period 1888–1913, constructed from tabulations of yields of primary crops in yearbooks published by Imperial Russia’s ministry of interior. This panel is meant to help identify the effect of temporary income shocks.

In the model, individuals living in the country of origin face an optimal stopping time problem, as in Rust (1987). Each period they observe the state of the economies in the country of origin and the country of destination, and form expectations about their future paths. They maximize expected lifetime utility by deciding whether they should migrate, or stay and face a similar dilemma in the next period. During recessions in the country of destination, migration is more (broadly-defined) costly. Migrants expect recessions to be short lived and may choose to postpone their migration,

---

2 Section 2.2.3 discusses alternative, more widely accepted views.

3 In practice it does so with little success; see discussion in section 6.1.
knowing that the option to migrate is kept and could be exercised in the near future as the recession subsides. Prospective migrants are heterogenous, unobservably distributed along a range of types over the stayers-movers scale, where “movers” expect to receive relatively more flow utility in the country of destination. Each district is characterized by a different mean of this distribution, such that districts with higher means have a potential to produce greater rates of migration. An extension to the benchmark model allows an individual’s type to fluctuate stochastically over time.

The dynamics is illustrated in Figure 3. During a recession in the country of destination, there is a decline in the rate of migration. But as the recession subsides, those who would have migrated absent the recession are still very likely to migrate. Some of them do so over the course of the coming years, thus increasing the number of migrants compared to a no-recession scenario, and offsetting the short-term effect of the income shock. A positive shock to past migration affects current migration through a level effect, in reducing the number of current prospective migrants, and through a composition effect, by removing a disproportionate number of “movers” and shifting current distribution towards more “stayers”. Both the level and the composition effects make a positive shock to past migration reduce the current number of migrants. For a given shock to past migration, the more inegalitarian the types distribution is, the stronger is the composition effect, and the greater is the effect on current migration probabilities. Therefore, the distribution of unobserved types is a key determinant of the dynamics of aggregate migration, and it ought to be estimated in order to evaluate the effects of income on migration.

The diffusion of networks is modeled as a process in which options for migration are stochastically disseminated to individuals. At any given period, a prospective migrant is either linked or unlinked. If unlinked, he cannot migrate; when linked, the dynamic migration problem described above applies. Linkage is irreversible. The probability to become linked at a given period is a function of recent migration from the individual’s district and from his province, as proxied by the incorporation of immigrants associations. Thus, the network effects are modeled as past migration increasing the linkage probabilities of geographically related prospective migrants. Pioneers are allowed, in the sense that linkage may occur even when there was no past migration. Throughout the migration movement, districts are advancing from being unexposed, with few linked individuals, to being saturated—a state in which almost all residents are already linked, and more migration no longer increases linkage rates.

The estimation follows a Maximum Likelihood procedure. An instance is a cohort-district vector of yearly counts of migrants. The likelihood of observing an instance, given the age and the realizations of business cycle fluctuations, is calculated while integrating over the unobserved distribution of types and of linkages. Since the cohorts in each district had been exposed to migration prior to the period for which direct migration data is observed, the initial distributions of types, as of the first observed period, are changed by past migration; the estimation controls for this initial conditions problem by correcting the initial distributions of types according to the migration probabilities in years prior to the observed data. The cohort-district likelihoods are thus calculated as
a function of the district mean type. Aggregating at the district level, the district mean types are either integrated out as random-effects with a known distribution, or directly estimated as district fixed-effects without distributional assumptions. In the extension in which individuals’ types can fluctuate around an individual-specific mean, the unobserved heterogeneity is compounded by a serially-correlated stochastic state variable. To overcome the challenge of tractability in integrating over the very large number of possible paths of this unobserved variable, the estimation uses a Simulated Maximum Likelihood approach, by which probabilities are computed for a random sample of possible paths.

The problem of separately identifying the long-run effects of income is not completely solved due to lack of direct panel or cross-section data on income at the district level. I show how, given such data, these effects would be identified. The currently available data do, however, enable to perform short-run analysis and simulate effects of transitory income shocks on migration. In simulations based on the benchmark estimation model, when the cohort aged 20 experiences a transitory negative income shock that is allowed to persist with gradual decay, on average its rate of migration decrease during the current year and during the next three years compared to a no-shock scenario. However, almost half of the would-be-migrants whose migration was avoided during the first four years would have migrated in one of the years until age 30. When simulating income shock with a “reset” (i.e., the business cycles proceeds in the second year as if income was on trend in the first year, such that the initial shock to migration probabilities extends through a single period only), almost 80 percent of the first year shock is offset by delayed migration by age 30. The conclusion is that transitory income shocks mostly affect the timing of migration, but have much smaller effect on whether any given individual would migrate or not during his lifetime.

I find the diffusion process to be a very strong predictor of the rates of migration. During the period of observed migration (FY 1900–1914), all districts had already been exposed to migration to some degree. Most are already saturated, while a minority are in the process of becoming saturated, some of them starting with under 50 percent linkage in FY 1900. By FY 1914, the diffusion of migration networks is complete and saturation is almost global. Thus, the period of the observed migration provides a view onto to the final stages of the diffusion process. I show that the identification of the parameters of this process is based on the correlation between the networks build up (as measured by the associations data) and the rate in which the rates of migration are accelerated during the observed period. Lower rates of prior exposure and higher rates of new linkage formation are associated with faster increase in the rates of migration during FY 1900–1914. This is a convergence process predicted by the diffusion model: districts that start the period with low linkage rates are becoming linked, and catch up with their migration potential; saturated districts are not increasing their rates of migration.

4 The problem of estimating a dynamic model with serially-correlated unobserved random variable was previously explored by Keane and Sauer (2009) and Gallant, Hong, and Khawaja (2010).

5 This means that I am unable to answer questions such as how would the rate of migration change had Russian income been permanently 10 percent greater than it actually was. See details in Sections 4.2.4 and 6.
By extrapolating backwards, to the period on which the data on associations exist but there are no direct measures of migration (1861–1899), the estimation enables to predict and outline the diffusion process from its very beginning. It shows that the linkage process was gradual and advanced spatially. Prior to the 1880s, only a handful of districts had been linked. Most were first exposed and became linked during the next two decades. This implies that the potential for full fledged mass migration was for a long period inhibited by the slow arrival of migration networks.

Can we learn from the case of the Jewish migration from the Pale of Settlement on the general questions of the Age of Mass Migration and the economics of migration in general? While peculiar in many ways, I argue that the Jewish population of the Pale of Settlement makes up a useful case study for mass migration. Its rates of migration were among the highest in Europe for several decades. This population was at least as large as any of the few identifiable ethnic or national groups that had experienced the highest yearly rates of U.S. immigration. Unlike the case for some other ethnicities, such as the Italians, the U.S. stood prominently above other destinations and clearly received the lion’s share. This conveniently enables to avoid complexities and model the process as a simple yet realistic one-to-one migration problem.

Jews resided in relatively uniform distribution across more than 230 districts in the Pale of Settlement, a wide and economically diverse area of over 1.2 million square kilometer (as large as the combined area of the German and the Habsburg Empires, and certainly greater than any other European country), that was nevertheless governed by similar rules and the same rulers and institutions. Russian statistical sources offer ample data on these districts and on their Jewish residents. Additionally, very few Jews lived in non-urban localities, almost all of them lived in towns inside the Pale. This facilitated the identification of their precise last place residence, compared to agricultural populations that came from myriads of hard-to-identify villages. Jews stood out with the lowest rates of return migration. This eliminates the problem of confounding temporary migration, which was likely much more sensitive to business cycles fluctuations, with permanent migration, the phenomenon of interest.

The paper proceeds as follows: Section 2 covers the historical and theoretical background, Section 3 describes the data, Section 4 explains the model and Section 5 the estimation procedure. The results are reported in section 6, and the final section concludes.