

The Dynamics of Mass Migration

Estimating the Effect of Income Differences on Migration
in a Dynamic Model of Discrete Choice with Diffusion^{*}

Yannay Spitzer[†]

Brown University

November 25, 2014

[PRELIMINARY AND INCOMPLETE]

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 Crown 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.

[†]Population Studies and Training Center, Brown University. 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, lag behind and took several decades to catch up and surpass the migration from the wealthier western European countries?

Using the evidence of the massive Jewish migration from Russia in the late nineteenth and the early twentieth, this paper revises the debate on the European pattern of transatlantic mass migration, and advances a dual explanation: (a) The sharp procyclicality 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

¹ For a comprehensive discussion of pre-WWI patterns of transatlantic mass migration, see Hatton and Williamson (1998).

had already migrated.

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 Eurenium (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 (1980) and Baines (1995), but lack of sufficient data and estimation methods has left the verdict on this hypothesis still pending.² Following Spitzer (2014), who 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.

This study is enabled by the creation of a data set of unprecedented span and resolution. 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 *landsmanshaftn*—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 needed 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

² Section 2.2.3 discusses alternative, more widely accepted views.

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, 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, since such within-individual time variation potentially affects the degree of persistence of temporary shock.

The dynamics are 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 unequal 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.³ 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.⁴ 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 decreases 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 a much smaller effect on the likelihood of any given individual migrating 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

³ 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).

⁴ 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.

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.

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.⁵

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

⁵ In principle, an extension to this model could incorporate the option to return-migrate, although it would be hard to explain within such model why conditional on having migrated, some groups, such as the Italians, tended to return whereas Jews did not. The model also does not capture the household problem, ignoring the question why Jews migrated en famille.

are reported in section 6, and the final section concludes.

2 Background

2.1 The Economics of World Mass Migration

2.1.1 The Pre-WWI Transatlantic Migration

From Waterloo to Sarajevo, European migration to the New World had increased from a trickle to a tide, reaching unparalleled dimensions until it was abruptly brought to a halt by the outbreak of World War I. The United States was the primary destination. From yearly rates of 40-80 thousand immigrants in the 1830s, the trend of U.S. bound immigration rose up, reaching hundreds of thousands during the second half of the nineteenth century, and almost a million a year between 1901 and 1914.⁶ This flow amounted to one of the greatest and most consequential events of cross-continental human movement.

This mass migration, whose rise coincided with the decline of the Atlantic slave trade and the European indentured transatlantic migration during the earlier part of the 19th Century, is typically attributed to a significant decline in transportation costs through the first and second third of the nineteenth century, which made migration more affordable compared to the benefits in terms of expected lifetime gain in income. It also improved the passage of information, which reduced the uncertainties of the move. Other than that, while some of the migrants may have been driven out by political and social upheavals or oppression, the dominant view is that the overwhelming majority of migrants moved to opportunity, being attracted by the prospects of higher real wages in the United States and escaping European poverty, demographic pressures, negative effects of industrialization, and in some cases famine (Hatton and Williamson 1998, p. 12).

An important factor that enabled that period to become a stage to a largely economic migration was that the movement of people was for the most part unobstructed. Few major wars took place throughout it and the restrictions on entrance to the United States were as lax as they could be. With few exceptions, any able bodied European who was capable of undertaking the voyage could become an immigrant. More than in any time since, people were free to react to economic incentives in choosing whether to migrate or not. Additionally, since much of the migration taking place at this period was well documented, it remains the primary event which can inform researches on the way economic conditions affect migration.

Events that followed after 1914 imposed a drastic change in this regard. The interruptions caused by World War I brought the transatlantic migration to an almost complete cessation, while the immigration act of 1924 terminated the liberal admission policy and drastically limited the immigration from the south and the east of Europe. After a second long interruption during the Great Depression and World War II, the levels of flows of world mass migration gradually gained back their former absolute (although not relative) magnitude through the second half of the twen-

⁶ For the yearly figures see Ferenczi and Willcox (1929, Table IV).

tieth century. But it took place under heavy restrictions posed by state policies, discouraging many would-be emigrants and forcing many others to go under the radar and move illegally and undocumented (Hatton and Williamson 2008, parts II and III).

2.1.2 Income and Migration: A Puzzle

In addition to the general rising trend in its levels up until World War I, two major features of the transatlantic mass migration were noticed: (a) Within countries, over time, it was always very volatile, and apparently very sensitive to temporal changes in income levels; and (b) while there was great geographical variation in rates of emigration within time, across different sending countries, this variation did not always seem to correlate clearly with levels of income among these countries.

U.S. bound immigration fell dramatically and almost immediately with each of the major recessions of the period (see Figure 1a). During the Long Depression that began in 1873, it went down from a peak of 437,304 in 1872 to a trough of 130,507 in 1877, a more than 70 percent drop. It rebounded to 67 percent *above* the previous peak level by 1881, almost 5 times over the trough. The peak-to-trough drop was 52 percent during the Depression of 1882-1885 and almost 63 percent during the depression ensued by the Panic of 1893. The financial crisis caused by the Panic of 1907 brought about a 69 percent free-fall in immigration by 1908, with an almost equally rapid recovery. Indeed, taking the time-series data on immigration at face-value, the straight-forward conclusion is that migration is highly sensitive to relatively small changes in the levels of income.

Early quantitative studies picked up on this pattern, and by examining data within different countries were mainly occupied with the *push vs. pull* question, of which economic conditions mattered more, those in the country of origin or in the country of destination. Among them were Jerome (1926), Dorothy Thomas (1941), and the influential study by Brinley Thomas (1954), who argued that the transatlantic migration in itself was a cause of business cycle fluctuations and to the fact that American and British business cycles at that time were negatively correlated. As economic historians began to study the issue using econometric methods, many of them followed the early studies by performing time-series estimations over various countries and specifications. They typically found that short term changes in income or unemployment cause very large changes in the levels of migration.⁷

As opposed to the evidence on high sensitivity to migration from the time-series analyses, the cross-sectional and panel-data evidence is weak at best. Figure 2 shows how total emigration rates correlated with levels of real wages in several European home countries over four decades. At the end of the period, during the first 14 years of the twentieth century, the relation is more or less as one would expect: poorer countries, such as Italy, Spain and Portugal, were sending more emigrants than wealthier countries in the north-west of the European continents. But going back

⁷ For surveys see Gould (1979) and Hatton (2010).

in time, this relation becomes weaker and even reverses—during the 1870s, the poorer countries of the south of Europe had hardly entered the migration cycle, while the United Kingdom, Europe’s wealthiest, was almost at the lead in emigration rates as well. Figure 2 does not report data on eastern European countries such as the Russian and Habsburg Empires, but including them would have strengthened the perplexing evidence, as they were both relatively poor and later-comers to transatlantic emigration.

There is a limit to the power of cross-country based inference, and so “To know more about European migration we must disaggregate” by going down to the regional level.⁸ However, evidence from within-country comparisons, while often showing the expected negative correlation between measures of standards of living and emigration, suggest that there were large degrees of variations even between similar regions, and that economic conditions do little to explain much of them. This point was made by Gould (1980), who speculated that the within-country variation in emigration might be declining over time, and by Baines (1994, 1995), who argued that often it does not. Micro-level studies by Simone Wegge, such as Wegge (1998), showed that even across adjacent German villages there were great differences in the scale and character of emigration.

2.2 Short- vs. Long-Run Effects of Income

2.2.1 A Problem of Identification

Most attempts to econometrically estimate the determinants of migration used an OLS estimation in which levels of migration were regressed on levels of income and other variables. When migration was regressed in time-series regressions on the levels of income (either the difference between the origin and the destination or the two levels of income separately), and business cycle variables were not included as control variables, the effects of income proved to be very strong. But when business cycle variables, such as rates of unemployment or the deviation of income from a long-run trend, were added to the regressions they typically picked up the entire explanatory power of income, and the effect of income absent the business cycles showed weak or insignificant. This problem was surveyed by Gould (1979), who expressed disappointment from the lack of capacity of the empirical literature to differentiate properly between the short-run and the long-run effects of income.

Sixteen years later, Hatton (1995, p. 408) reported no significant further advance in the literature, and proposed a new model of micro-foundations, in which the components of the individual decision to migrate are taken into account, including how expectations are formed regarding future benefits of migration. It rationalized the strong reaction to business cycles as the effect of increased risk of unemployment, which makes migration during a recession in the country of destination more costly. It incorporated the friends and relatives effects by adding stock and lags of previous migrants, and showed how the option to wait could be captured in the micro-model that motivated the

⁸ Baines (1995, p. 25).

estimation. The micro-foundations were linked to a macro-level reduced form econometric model that effectively looked very similar to previous estimations: it regressed migration on income differences and business cycle variables while controlling for other variables such as past migration. This model was used in a time series estimation of U.K. emigration, and in Hatton and Williamson (1998, Ch. 4) it was also applied to the case of Scandinavian emigration. Boustan (2007) and Bohlin and Eurenus (2010) used it in studying the Jewish migration from Russia and the Swedish emigration.

In what follows I show formally the identification problem in the time-series estimation. Let w_t^0 and w_t^1 be the log of real income levels in period t in the countries of origin and destination, respectively. Let d_t^0 and d_t^1 be the deviations of log real income from a linear trend at period t , representing the states of the business cycles. Finally, the rate of migration at period t (in log per capita terms, for example), will be denoted by m_t . The time series OLS estimation has the following specification:

$$m_t = \alpha + \beta \Delta w_t + \delta_0 d_t^0 + \delta_1 d_t^1 + \gamma z_t + \epsilon_t \quad (1)$$

where $\Delta w_t = w_t^1 - w_t^0$ is the real wage gap and z_t is a vector of control variables representing characteristics such as stock and lagged migration, demographic variables, etc. Since the business cycle variables are defined as the deviations from long run income trends, typically linear trends, the real income gap variable could be written as

$$\begin{aligned} \Delta w_t &= w_t^1 - w_t^0 = (\tilde{w}_t^1 + d_t^1) - (\tilde{w}_t^0 + d_t^0) \\ &= \Delta \rho_a + \Delta \rho_b t + d_t^1 - d_t^0 \end{aligned} \quad (2)$$

where \tilde{w}_t^1 and \tilde{w}_t^0 are the income trends, and $\Delta \rho_a$ and $\Delta \rho_b$ are the differences in the intercepts and the time coefficients of the linear time trends. Not surprisingly, this representation shows that the level of real income difference is nothing but a linear combination of time and of business cycle shocks. Effectively, when both the difference in real income levels and the business cycle variables are included in the equation, it becomes econometrically equivalent to running a regression on the business cycle variables and on time.⁹

The only source of identification for the effects of the levels of income on migration thus stems from the advance of time. Keeping the business cycle variables fixed, there is no additional information in the levels of real income other than what is predicted by the time trend. This is a very poor source of identification and it can lead to somewhat perverse results. For example, if there is

⁹ This does not get better when separating the difference in income levels and regressing on the two respective income levels; the four variables of income levels and business cycles are a linear combination of one another, resulting in a perfect multi-collinearity in the regression.

convergence of income levels over time while the rates of migration are growing (as was the case in Russia, that was way behind American income at that time but with a marginally faster growth, as well as in Italy (Moretti 1999)), controlling for business cycles, the effect of real income difference is likely to be estimated negative. This problem of identification was driving the phenomenon pointed out by Gould (1979) and it is endemic to all time-series specification. The model in Hatton (1995) attempted to identify β , the parameter governing the long-run relation between income and migration also from variations in the business cycle shocks d_t ; this came at the cost of making a very strong calibration assumption that enabled interpreting the effect of the business cycle shocks in terms of changes to the utility from income.¹⁰

A few studies estimated the effects of income on migration using regional panel data. On the pre-WWI migration, Hatton and Williamson (1998) used decade data on 32 districts in Ireland and 69 provinces in Italy, and Snchez-Alonso (2000) used data on 49 Spanish provinces in two points in time. But with the coarse decade-or-longer time periods, there is no hope in learning about the effects of the short-run business cycle fluctuations in these cases. Recently, this was improved upon by Bohlin and Eurenium (2010), that employed yearly panel data on emigration from 20 districts in Sweden. On the late 20th century migration a few works employed panel data of country-to-country migration. Among them Karemera, Oguledo, and Davis (2000) with data on migration from 70 countries to North America, Clark, Hatton, and Williamson (2007), 81 countries to the U.S.), and Pedersen, Pytlikova, and Smith (2008) and Mayda (2009), that employed OECD data on *legal* migration from 129 and 79 countries (respectively) to 22 and 14 OECD countries.

2.2.2 Business Cycles and Timing

The very strong sensitivity of rates of migration to business cycle fluctuations is in itself a puzzling phenomenon, and it has recently been recognized as yet unresolved (Hatton 2010, p. 945). As deep as recessions may have been, they were short lived and their effects on the expected lifetime income was negligible compared to the sheer income differences between the origin and the destination. It is clear that there is something about recessions that makes it particularly undesirable to migrate as long as they take place. In Hatton (1995) and Hatton and Williamson (1998), this undesirability was interpreted as the effect of greater income uncertainty due to the greater chance to be unemployed. Indeed, immigrants are particularly exposed to small changes in unemployment rate as they are, almost by definition, unemployed upon arrival. If the rate of U.S. unemployment doubles during a recession from 3 percent to 6 percent, then a prospective immigrant would have to compete with twice as many job-seekers. Since many of the immigrants jobs at that time were short-termed, a recession immigrant should have expected to be employed through fewer days per year, or to settle for lower-paid jobs. It may be harder for relatives and friends, whose resources constraints may be temporarily tightened, to support him upon arrival. One way or the other, there is something

¹⁰ More details on the challenge of separately identifying the short-term from the long-run effects of income are discussed in Section 4.2.4.

more costly, in terms of broadly defined costs, in recession immigration.

However, the traditional OLS estimation that controls for both the levels of income and for the business cycle fluctuations implies that, first, a recession does not change *at all* the expectations of future income; and second, that *none* of the temporary decline in migration is later compensated by delayed migration, prospective immigrants who would have migrated absent the recession and were potentially still looking up for migration as the recession subsides. It seems plausible, on the contrary, that to some degree both of these implications are wrong: Since recessionary shocks do have some persistence, a temporary shock informs the prospective migrant not only that migration today is more costly, but also that in the next few years he should expect somewhat lower levels of income. Ignoring this effect by attributing the entire change in migration to the contemporary shock and none of it to the change in future expectations would lead to somewhat over-estimating the effect of the short-run changes in income while under-estimating the effects of the long-run differences. The estimation model I propose captures the effect of temporary shocks on future income and corrects for this mis-specification.

The second case is exemplified by Figure 3. The upper diagram shows a schematic progression of the levels of migration absent a recessionary shock. The lower diagram adds in a recession occurring around period t , that drives down significantly the rate of migration during the time of the recession. The area plotted with white dots represents the mass of immigrants that avoided migration due to the recession. The recession ends around period $t + 1$, and this mass of would-be immigrants is still present among the pool of prospective migrants in the country of origin. In fact, since they were very recently apt to migration, they are very likely to still be more apt to it in periods $t + 1$ and later compared to an average individual in their demographic group. Hence, they may choose to migrate in one of the years following the recession, resulting in a somewhat larger rate of migration in those periods. This additional stream of delayed migration is represented by the area plotted with black dots.

By ignoring the case of delayed migration following a recession, the estimation does not allow for the temporary effect of recessions to be partly offset in the long run. The answer to the question *how many of the lost immigrants due to a recession will immigrate in the future*, is zero by assumption. It was speculated that some mechanism of this sort does take action,¹¹ but no empirical test has yet been attempted. If this happens, then the short-term effect of a recession may largely be offset by delayed migration, and the puzzle is solved: short-term income shocks mostly affect the timing, not the ultimate long-run level of migration. The implication is that past conditions matter in predicting probabilities of future migration, and that a dynamic model is necessary in order to capture the mechanism of delayed migration. The estimation model proposed in this paper is designed to capture this mechanism, and in section VI I show evidence that this is indeed what happens. For every two recession non-migrants, only one will remain forever in his home country. The other will migrate some time at a later period.

¹¹ See Baines (1995, p. 15); Hatton (2010, p. 945).

2.2.3 The Role of Migration Networks

Migrations take place in chains. It is an established fact in the literature that past migration is one of the best determinants of future migration. The reasons for this are fairly clear: migration is a costly endeavor that involves great uncertainties and the need to overcome liquidity constraints. Even when the pecuniary costs of migration are easily offset by the expected gains, a prospective migrant will find it almost impossible to raise resources by his own and migrate as a pioneer. Historically, there was a number of ways to overcome the liquidity and information constraints fundamental to the costly and risky transatlantic migration. The slave-trade solved it by using the institution of slavery to guarantee some of the surplus from migration to the trader. Similarly, indentured labor contracts, the most typical sort of European transatlantic migration during the eighteenth century,[REFERENCE] solved the problem by binding contracts that guaranteed an ex-post payment. As will be described later on, the Jewish migration suffered severely, at the beginning, from this problem, and it was solved eventually by the thickening of migration networks: previous immigrants who were closely related to prospective migrants in the old country supported their migration by funds, information, and material support. For most of the Jewish immigrants, this support was a necessary condition for migration.

The significance of migration networks in explaining the patterns of the European transatlantic migration was demonstrated by Gould (1980), who argued that the only way to explain why East- and South- European countries entered the scene of transatlantic migration only around the turn of the twentieth century and not decades earlier, is due to the slow progression of networks across space in Europe. Particular cases of chain migrations can only be explained by a chance event that had started them early on, such as the Italian sailor who had broken his leg during the journey and had to rest for a few months, and eventually became single-leggedly responsible for a stream of Italian migration to Wellington, New-Zealand, wholly originated from a single Calabrian town. Moretti (1999) showed that considering migration networks is crucial in the case of Italian emigration, and Wegge (1998) demonstrated how such networks were often very localized. In Hatton and Williamson (1998) the past stock and lags of migration flows were used as an important explanatory variable, with the interpretation that networks make migration less costly. On recent migration, Munshi (2003) provided evidence of how the exogenous growth of networks increases future migration in the case of Mexicans in the U.S.

If networks of previous immigrants do not only *facilitate* migration but virtually *enable* it, then the prediction should be that whatever affects migration, does so with interaction with the size of the existing networks. For example, a model that interprets migration as mere reduction in migration costs would imply that the same business cycle shock would have approximately the same effect on migration, almost regardless of the size of the network. On the other hand, treating instead the size of the network as a determinant of the number of individuals who have an *option* to migrate, would imply that a shock that occurs when the networks are well saturated will have a larger effect on migration than when the networks are at the stage of burgeoning or when they do not yet

exist.

I argue that the interpretation of networks as a necessary condition rather than a cost-reducing institutions is a far better first-order approximation to the way migration actually worked, and that this has meaningful predictions that could be incorporated within an estimation model and be tested empirically. All evidence suggest that the networks of migration were mainly based on personal encounters or small close-knitted communities and in most cases close family relations. As will be discussed later, that was particularly true for the Jewish migration.

Jewish migrants certainly benefited from the fact that there was a large population of Yiddish-speakers in New-York that could have supported services and amenities such as education, newspapers, and theater; but at the core, and as a first degree approximation, having a closely related person already there made the difference between a person who could have responded to economic and other incentives and migrate if he found it worthy, and an unlinked person who could not ascend the liquidity and information constraints.¹² Whether there were 400 thousand or 800 thousand Jews in New-York at a certain point in time probably mattered little for a would-be migrant that was linked anyway. But for an East-European Jew that gained a link in America thanks to a single relative or fellow hometown that could support his migration, that single person made the whole difference.

The interpretation of networks as institutions that create links for prospective migrants could offer a natural explanation for the geographical puzzle, of why poorer regions in the east and the south of Europe were late-comers to transatlantic migration: it happened so because it took time for migration networks to diffuse eastwards and southwards and to build up there. The Jewish and the Italian migrations turned from a thin trickle prior to the 1880s into a gushing flow around the turn of the century only because too few of the Jews and the Italians were linked at the early period, and two decades later the migration networks have reached near saturation. There is disagreement in the literature on this point. As against Gould (1980) and Baines (1995), who argued for the significance of spatial diffusion as a factor delaying migration, Hatton and Williamson (1998) have objected to the view that the timing of the onset of mass migration is explained by it. While the gradual growth of local networks was a crucial factor in delaying full fledged emigration, what determined the time in which the process had began in each country were local economic and demographic conditions. These were mainly the advent of industrialization, urbanization, and following Easterlin (1961) also rising population pressures caused by the demographic transition. The claim that a slow spatial diffusion was a determinant of the timing of mass migration was dismissed by stating that “[...] it offers few empirical predictions and says nothing about why emigration rates eventually decline.” (p. 15). I disagree with the first part of this statement and believe that the second part is irrelevant

¹² Furthermore, it is not clear that the overall number of migrants from the same country, but not from the same region, does not run into diminishing or even decreasing returns to scale. Jews have specialized in particular occupations and were heavily clustered in the city of New-York. An argument could be made that beyond a certain level the greater economic competition emanating from more fellow-countrymen in the same market outweighed the benefits.

for judging the matter. The spatial-diffusion explanation can and should deserve a fair chance to be approved or dismissed based on empirical evidence.

The estimation model I use explicitly treats migration networks as generators of migration ‘options’. I show in sections IV and VI, that the data on the case of the Jewish migration appears consistent with this interpretation. Pioneer migrations could still take place, as migration options can be produced also without networks, it just occurs so more rarely. At the other end of the scale, migration networks can reach saturation, a point by which almost all of the home-region population is linked to previous immigrants and have the option to migrate. In between, the thickness of the network plays an important role in determining the onset and the scale of mass migration.

2.3 The Jewish Migration from the Pale of Settlement

2.3.1 Jews in Russia: An Overview

At the turn of the century, the Russian Empire was home to some 5.3 million Jews, approximately half of world Jewry. Almost all of them, 94 percent, were residing in a restricted territory known as the *Pale of Settlement*, comprised the 25 western provinces of the Russian Empire. Residence of Jews beyond the Pale was severely restricted by a set laws and statutes enacted during the decades following the Polish Partitions and the Russian takeover of the lion’s share of the Polish-Lithuanian Commonwealth (Klier 1986). Within the Pale, the Jewish population was typically concentrated in small provincial market towns, known as *shtetl*, in which Jews made up the majority of the population (Rowland 1986a). An ethnic minority of some 9 percent of the total population (as of 1897), Jews have specialized in certain occupational sectors: almost none were farmers, and about a third was employed in manufacturing, where they were over-represented compared to other ethnicities. Another third was employed in trade and commerce, an occupational niche in which they outnumbered the rest of the population in absolute terms (Kahan 1986a; Rubinow 1907).¹³

Under the Tsars, the Jewish population had gone through a very rapid population growth, according to estimates as much as five-fold during the nineteenth century (Stampfer 1989). By the end of the century it was generally poverty-stricken, and for the most part on the losing side of the transformations brought about by the advent of Russian and Polish industrialization (Kahan 1986a).¹⁴ There was some regional variation in this respect. It appears that the conditions in the South-West and the region known as *New-Russia*, comprising of the southern provinces bordering with the Black Sea, were better, with higher real wages and improved standards of living.¹⁵

¹³ On the Pale of Settlement Jews as a prime example of ethnic minorities specializing as middlemen see Slezkine (2004). For an analysis of Jewish occupations based on detailed data from the 1897 census, see yannayspitzer.net/2012/09/30/jewish-occupations-in-the-pale-of-settlement.

¹⁴ For a classic account of the conditions of Jewish laborers in the North-West see Mendelsohn (1970). On the the social-economic conditions of the Jews in relation to the migration to America see Lederhendler (2009).

¹⁵ Rubinow (1907) has provided summaries of some spotty regional wage data from the 1898-1899 survey of the

The relations between the Jewish population and the Russian Tsars, their bureaucracy, the Intelligentsia, and the people were complex and at times tumultuous.¹⁶ Prior to the Polish-Partitions Jews were banned from residence within Russia, and the presence of masses of Jewish communities within the newly acquired territories had forced a complex process of accommodation with the *Jewish Question*. It was hardly ever smooth and it has not nearly seen completion by the end of the Imperial period. Main issues at stake were the economic activities of the Jews, often seen harmful and in need of correction; their religious "fanaticism"; their education; their rights of occupation and residence; their duties, in terms of taxation and military service; and their political rights, mainly as far as participation in local politics and the legitimacy of Jewish autonomous institutions.¹⁷ Many Jews felt threatened by constant attempts of an absolutist monarchy to encroach upon their communal autonomy and their traditional ways of life. Never having granted equality of rights to all its citizens before its demise, the Russian Empire had sets of discriminatory rules, in which different classes, ethnicities and religious minorities were designated with different rights, duties and bans. While other groups have suffered from these circumstances as well, the Jews never felt particularly favored by them.

2.3.2 Jews Leaving Russia: An Overview

The year 1881, particularly the crisis that followed the assassination of the relatively liberal-minded Tsar Alexander II, and the ascendance to throne of his more reactionary son Tsar Alexander III, are often considered a revolutionary "turning point" in Jewish History.¹⁸ It marked a transition to modern East European Judaism, the emergence of the Zionist movement, of Jewish Socialism, and not least, of mass overseas emigration from Russia. A wave of *pogroms*, anti-Jewish riots, broke out that year in the southern city of Elizavetgrad and spread out to many other towns, mainly in New-Russia and the South-West. A year later, a similar wave broke out again in the same regions. In response came the notorious *May Laws* of 1882, that further restricted the rights of residence and occupation of Russian Jews, and was followed by similar acts of legislation in subsequent years.¹⁹ The prevalent view that there has been some orchestration of these pogroms from the top was dismissed in a number of revisionist studies from the past generation,²⁰ and indeed, compared to similar events yet to come during the twentieth century the number of casualties was not as high. Nevertheless, there is little question that the pogroms and the anti-Jewish legislative surge have

JCA (Association 1906, see), but no systematic data exists. In section III I report data from the 1897 Russian census shedding new light on the differences between the Pale's regions.

¹⁶ Dubnow (1916) is the fundamental study of this topic, notwithstanding many revisions of views since its publication. See also Baron (1976).

¹⁷ Such issues were comprehensively covered by Klier (1995) for the reign of Alexander II (1855-1881). On Jews in the Russian army see Petrovsky-Shtern (2008).

¹⁸ This view was expounded by Frankel (1981), and many studies have used the year 1881 as a beginning or end periodization point. For a more gradualist approach see Nathans (2002).

¹⁹ For a compilation of a list of Russian laws specifically discriminating against Jews reported by the Weber-Kempster report (Weber and Kempster 1892) see Johnpoll (1995). For a similar British report see Committee (1890).

²⁰ See Rogger (1986), Aronson (1990), Klier (1992), and Klier (2011).

contributed to the sense that the conditions of the Jews in Russia were such that matters ought to change and radical solutions should be sought. Emigration was one of the solutions that were found on the shelf.²¹

Whether 1881 was indeed a radical turning point in emigration growth remains debatable,²² but to be sure, that was the first year in which Jewish Russian Immigration to the U.S. had crossed the mark of 10 thousand yearly entries. Figure 4 shows the levels of Jewish migration in the period 1880-1913, plotted against the American and the Russian business cycles.²³ During the 1880s there was a steady rise, although not monotonic, in the level of migration. The year 1891 saw a sharp peak, corresponding with a severe harvest failure in Russia, and also with an episode of mass deportations of Jews from Moscow. The following years, during which the U.S. economy was hit by the 1893 Panic and a prolonged recession, and Russia benefited from a period of growth brought by a spurt of industrialization, migration came back to the 1880s levels.

Jewish Russian migration gained momentum around the turn of the century, and peaked during the years 1903-1906, a period of economic and political turmoil in Russia: It saw the Russo-Japanese War, the 1905 Revolution and a new wave of pogroms starting with the shocking events of the *Kishinev Pogrom* of 1903 and culminating with an explosion of hundreds of pogroms in 1905, during the week following the issue of the *October Manifesto* by Tsar Nikolai II.²⁴ Concurrent with the Russian recovery and the U.S. Panic of 1907, the levels of Jewish migration have fell down dramatically to a low point in 1908, gradually recovering up until World War I, which has brought an almost complete cessation in transatlantic travel.

The economics of the Jewish Russian migration was studied by early works such as Rubinow (1907), Brutskus (1909), Hersch (1913), Joseph (1914), Ruppin (1934), and multiple studies by Jacob Lestschinsky, such as Lestschinsky (1961). A comprehensive paper by Simon Kuznets (1975) has summarized and added data and analysis and is still the most important study on the topic. The main quantitative features of the Jewish Russian migration are the following: Over three-quarters

²¹ Following the pogroms thousands of dislocated refugees have escaped the southern provinces and concentrated around the town of Brody, across the border in Austrian Galicia. Jewish organizations such as the Paris based *Alliance Israélite Universelle* tried to take charge of this crisis. After more than a year, during which the number of refugees had peaked to over 10 thousand at a certain point, they were resettled - many back in their home towns, some in various European countries, and others made their way to the United States (Szajkowski 1942). This event could stand as an example of a direct link between pogroms and emigration. In section III I show that there is little evidence that this flight of refugees had jump-started a strong flow of emigrants from the same region in the following years.

²² Kuznets (1975) estimated the number of Jewish immigrants from the Russian empire during the 1870s at 15-20 thousand, as opposed to 139.5 thousand during the 1880s. Diner (1995) claimed the migration had been building up since the 1860s and 1870s, and Alroey (2008) reported figures suggesting that Kuznets had under-estimated the volume of earlier migration.

²³ Migration was measured in fiscal years, starting July 1 of the previous year. In this paper I shift migration years going from July 1 of the *current* year to June 30th of the *next* year, so as to correspond with the years of the income series with a lag. E.g., The year 1880 stands here for the period July 1st, 1880 - June 30th 1881, etc. The migration data in the figure goes from July 1st, 1880 to June 30th 1914.

²⁴ The most detailed report of the 1903-1906 pogroms is found in Motzkin (1910), and it was used for the construction of the pogroms data set (see section III). For retrospective studies see Lambroza (1992) and the unpublished dissertation Lambroza (1981).

of the Jewish Russian emigration, 1.5 million, arrived to the United-States, while an additional half-million arrived from other East-European countries. Russian Jews had higher rates of migration compared to other East-European Jews, but not dramatically so. Compared to other ethnicities in the Russian empire Jews were highly over-represented among Russian immigrants, and they tended to have higher rates of non-labor-force participants among them (women, children, and elderly people), and laborers and manufacturing workers were over-represented among the immigrants compared to the total Jewish population in Russia. The rates of Jewish return migration were low compared to other ethnicities.²⁵ Most of them had their tickets pre-paid by relatives and friends in America, and in any case they were almost exclusively linked to a friend or a relative who had already migrated. Kuznets could not provide hard evidence on the specific geographic origins of Jews within Russia, but accepted the common view that Lithuanian Jews were over-represented. He believed that the pogroms and the persecution had some role in driving out Jews from Russia, but that it was not a crucial one, and speculated that the 1903-1906 wave of pogroms may have raised the share of migrants from the south, but not up to that of the migrants from Lithuania.

Since Kuznets, a number of economic and quantitative studies have shed further light on the questions of the geographic origins and the effects of the pogroms. Stampfer (1986) used data on the geographic origins of New-York *landsmanshaftn*, similar to the data described in section III, and showed that Lithuania was over-represented whereas New-Russia was under-represented. A study by Perlmann (2006) measured district-level migration using two samples of Ellis-Island arrival records, and has similarly found stronger propensity to migrate from Lithuania and the adjacent Polish districts. However, Alroey (2008, Table 4c) using data on applications for support in migration submitted to the *JCA*, has found an almost perfectly proportional representation across regions.²⁶ An econometric time-series study by Boustan (2007), using the model in Hatton (1995) while allowing for pogrom-effects for the years in which pogroms took place, found that there has been some increase in migration due to pogroms, but that it amounted to no more than a few tens of thousands of additional migrations.²⁷

The current paper adds over data produced in previous studies on the geographic origins of the Jewish migration in several ways. First, as opposed to previous time-series or cross-section data, it is a panel data, measuring yearly migration from almost all districts over 15 years, and thus it enables

²⁵ Sarna (1981) claimed that Jewish return migration was more prevalent than has been thought, but provided no quantitative evidence. Gould (1980, Table 3) showed that it was the *lowest* of all ethnicities. Recent evidence on return migration by Bandiera, Rasul, and Viarengo (2013), as well as indirect evidence by Abramitzky, Boustan, and Eriksson (2012), indicate that Russia stood out as the country with the lowest rates of return migration, which given the large share of Jews among Russian immigrants is consistent with Gould's figures.

²⁶ Alroey explained the discrepancy with Stampfer's findings in that the applications were by prospective migrants to all destinations and not just to the U.S.. However, As shown in his own work, migrants to other significant destinations such as South-Africa and Argentina were also disproportionately coming from Lithuania and Poland. Alternatively, part of this discrepancy could be explained by the fact that Alroey's applications data are mainly from the later years of the period, in which, as I show later, the geographic composition of migration was somewhat more balanced.

²⁷ Studies on other destinations, such as London (Godley 2001) and Ireland (Grda 2006) also found that Lithuania was over represented. In the case of Ireland, the origins of Jewish immigrants were narrowed down to a few Lithuanian districts.

to see the different dynamics of migration in different regions. Second, it uses a sample larger than previous studies, that is not very far from the entire population of U.S. Russian immigrants at that time. Third, the *landsmanshaftn* data enables to observe regional patterns in migration that go as further back as the 1860s, a period on which almost no quantitative evidence has yet been supplied in this regard.

3 Data

I construct a new and unique panel data measuring the number of Jewish immigrants from the Russian Empire over the period 1899-1914, with background characteristics pertaining to their places of origin. I aggregate the data across three dimensions to create measures of migration for each year, district, and cohort, over 15 years, 210 districts and 49 cohorts. The various sources I use to construct the data and the way they were used are briefly described below.

3.1 Ellis Island Arrival Records

Since the 1819 *Manifest of Immigration Act*, every ship entering a United States port was required to submit manifests with lists of all passengers aboard the vessel. At the beginning only the basic details, such as name, sex, age, and occupation, were required. Over time, the lists have expanded to include more details, among them the nationality of each passenger, based on the country of origin. Since 1899 it became customary, and since 1903 it was required, to report the ethnicity and the last place of residence as well (Weil (2000); Perlmann (2001)). In 1892, Ellis Island had began operation, replacing the older Castle Garden station to become the nation's largest facility for processing arriving immigrants. The overwhelming majority of Russian Jewish immigrants have entered the U.S. through Ellis Island ((Kuznets 1975)), and their details were recorded on the ship manifests.

The immigration data I use are based on those ship manifests submitted by shipping companies to the Bureau of Immigration in Ellis Island, in which the personal details of all immigrants arriving in the facility since 1892 were recorded.²⁸ While passenger ships manifests have long been used as a source in the study of immigration,²⁹ the records were only recently coded onto a machine readable file by the Church of Jesus Christ of Latter-Day Saints. The available data on all East-European immigrants in the years 1892-1924, more than 5.7 million individual records of migration, were given to me for research purpose by the *Statue of Liberty - Ellis Island Foundation*.³⁰ Among them, there are 2.33 million immigration records of passengers coming from the Russian Empire.

The first challenge pertaining to this data is to identify which of the passengers were Jewish. The problem is that even as the identification of Jews as a distinct ethnic group was required, too many Jews have gone unrecorded as such.³¹ Fortunately, I find that poor identification was rare, ship manifests either identified Jews, in which case they did it very well in almost all cases, or they did

²⁸ Unfortunately, very few records were left from the first 5 years of operation, due to the fire that had burned the entire immigration complex down in June 1897.

²⁹ Perlmann (2006) has used two samples of the original manifests to study the geographic origins of Russian Jewish immigrants.

³⁰ Another study, and to the best of my knowledge the only one, that has made use of the coded Ellis Island data is Bandiera, Rasul, and Viarengo (2013).

³¹ I estimate that as many as 25 percent of the Jews in the file were not tagged as Jews. This is partly due to unsystematic transcription of this field from the manifests, and in earlier years mainly due to partial recording of ethnicity on the manifests.

not identify Jews at all. The manifests that did identify Jews provide an Archimedean point to identify Jews systematically. I developed a simple algorithm that predicts whether each passenger was Jewish or not based on his or her first and last names. As a first stage, it uses the manifests that identified Jews to assign a measure of Jewishness to each first name and last name, as well as to their Soundex groups. At the second stage, it predicts whether each passenger is likely to be Jewish based on his or her first and last name. This algorithm yields very few false positives (i.e., cases in which a non-Jew is mistakenly identified as a Jew), while tagging most Jews as Jews.

The second challenge is to determine the last place of residence reported by each passenger and link it to an actual town in the Pale of Settlement. This identification faces daunting difficulties: the towns typically had Slavic names; they were reported by Yiddish-speaking passengers; handwritten by a German, British, or Dutch clerk on behalf of the steamship company; and finally a century later, deciphered and coded onto a file by a volunteer who in most cases did not know the geography of the Pale.

The strategy to address this problem is to tailor-fit a text condition for each and every town, that will match passengers based on the text of the "last place of residence" field, while taking into account the following difficulties: (a) phonetic variations and errors (e.g., *Kishineff*, instead of *Kishinev*); (b) graphic errors (e.g., "H" replaced for "K", as in *Hishinev*); (c) different towns with similar names (e.g., *Brichany* and *Brzeziny*); and (d) towns with multiple names or various pronunciation of the same name (e.g., *Warschau* [German], *Varshava* [Russian], and *Varshe* [Yiddish]).

At this point, the procedure has identified immigrants coming from the 450 largest Jewish communities, covering more than 3 million Jewish residents as of 1897. The effective coverage is surely higher than that, since many Jews coming from small townlets tended to report the nearby district town. Of the 2.33 million Russian immigrants in the file, 1.9 million have reported a potentially informative last place of residence; 779,286 of which I identify as Jews; 602,144 of which have arrived during the fiscal years 1899-1913; and to 295,626 of them I was able to link a particular town in the Pale of Settlement. The town-based identified migrations are aggregated at the district level, by years of birth, and by year of migration, to form the cohort-district-yearly measures of migration used in the econometric analysis.

3.2 The 1897 Russian Census

The 1897 Russian Census was the only general census conducted prior to the Russian Revolution, and is renowned for its relatively high quality (Clem 1986). While summary tables and individual figures were often cited from it, there remains a vast wealth of thousands of detailed tabulations, mostly at the local district level, that have not been coded or utilized in economic studies as of yet. Since many of the tables were cross-tabulated by ethnicities and religions, this census is the best available data on any Jewish population prior to the formation of the state of Israel in 1948,

capturing as much as half of world Jewry at that time.

First, a special volume within the census publications enables to map more than 85 percent of the Jewish population up to the level of the locality (Tsentralnyi Statisticheskii Komitet 1905). It lists each and every locality in the Russian Empire in which there were more than 500 inhabitants, and for each recorded locality it lists the populations of the religious minorities that comprised more than 10 percent of the total population. Since Jews typically lived in small provincial market towns, known as *shtetl* (a townlet, *shtetlach* in plural), in which they formed a majority of the population,³² this volume enables an almost complete mapping of the Jewish population in the Pale down to the level of the locality.

The *shtetlach* data set was generated from this volume by coding each and every town in European Russia in which a Jewish community was listed. A graphic representation of these data could be seen in the map in Figure 5. It is used to identify the towns that were to be linked to the Ellis Island arrival records, and it also links each town to its pertaining district, allowing the district-level aggregation of the immigration data.

The census publications contain a series of *guberniia*- (province) level volumes, one for each of the empire's provinces. For each of the 60 provinces of European Russia and the Kingdom of Poland, which include the 25 provinces of the Pale, I coded cross tabulations of ethnicities, age groups, literacy rates, and occupations. I use the tabulations of age groups for the Jewish population in each district to interpolate the size of each cohort in each district as of 1897.³³ I use occupational data, enumerating the number of Jews employed in each trade of a list of 65 different occupations, to calculate a district-level measure of the ratio of Jews employed in commerce and trade to the number of Jews employed in manufacturing. Since no systematic wage data exists, the commerce-manufacturing ratio uses as a proxy for the level of income of the Jewish population in each district.³⁴

3.3 Additional Sources

The map on figure ?? reports pogroms that took place during the violent wave of 1903-1906. The data was collected from two sources. Motzkin (1910) is a detailed report including chronologies of many major events, along side lists of minor events. It was completed by another, less detailed list published in the 1906/7 *American Jewish Year Book* (Szold 1906), and altogether the pogroms data includes 388 individual towns whose Jewish communities were known to have been hit at least once. According to Lambroza (1981), who collected information from additional sources,³⁵ these

³² On the definition of *shtetl* see Klier (2000). On the patterns of Jewish settlement in the Pale see Rowland (1986b)

³³ The interpolation uses polynomial splines to back out yearly cohorts from the 10-years age groups.

³⁴ On the correlation between this ratio and the standards of living of the Jews in the Pale see Rubinow (1907), Kuznets (1975), and Kahan (1986b).

³⁵ Unfortunately, in a personal communication with Shlomo Lambroza, I was told that the file generated for his doctoral dissertation and was coded on punch-cards has been lost.

two lists are close to comprehensive. What shows clearly in the map, is that the pogroms took place mainly in the South-West and in the southern provinces of New-Russia, and that Lithuania and Poland were largely spared.

The *landsmanshaftn* data is used to measure the evolution of district-level migration networks over time. A *landsmanshaft* is a generic name for hometown-based associations, prevalent in New-York and other large cities in the U.S. since the time of the migration and active well into the second half of the twentieth century. While in many historical cases of mass migration it was customary for immigrants who came from a particular region to form associations of mutual benefit or other purposes in the new country, the extent to which that was done by East European Jewish immigrants in New-York was unprecedented.³⁶ A survey conducted for the 1919-1920 American Jewish Year Book has counted over 5,000 Jewish organizations with over 1 million memberships, of which 2,421 were "fraternal orders and mutual benefits associations" with 574,163 memberships.³⁷ The *landsmanshaftn* were the main way of Jewish immigrants to continue the operation of some of the social and economic roles previously assumed by the town-based communities in the home country, but at the same time they were adapted to provide mutual support in the new circumstances in America.³⁸ One of their most important roles was to provide social and material support for recent immigrants from the same town.

I use a list of 3,014 hometown-based associations that were incorporated in the New-York County court during the period 1848-1920. As general a rule, the name of the pertaining hometown appears as a part of the name of the association, so that in most cases it is straight forward to link the associations to their hometowns. The court records note the year of incorporation. When immigrants from a particular town had incorporated an association in a particular year, I take it as an indication that around that time the network supporting immigrants from the respective town had thickened.³⁹

Figure ?? shows how a weighted measure of recently incorporated associations per-capita has evolved throughout the period. It reveals a few patterns on which previously only guesses could be made. First, it shows how the immigrants from the South-West and from New-Russia have entered the cycle of migration in an almost two decades delay, and that once they did that the rate of growth of their networks was not particularly rapid. Furthermore, the pioneers of the Jewish Russian migration were the Jews living in the provinces of Poland, followed by the Lithuanians. Once the Lithuanians joined in, the rate of growth of their migration networks has outpaced that of Lithuania. This is consistent both with the observation that Lithuanian Jews have suffered from worse standards of living, and with the view that the gradual pace in which migration networks

³⁶ In a comparative study of immigrants hometown-based associations, Moya (2005) has noted that the Yiddish term *landsmanshaft* became an accepted synonym for hometown associations in the historical literature.

³⁷ See Schneiderman (1919, p. 303).

³⁸ See Soyer (1997). For a case study on the associations formed by the town of Proskurov see Milamed (1986).

³⁹ Stampfer (1986) has used a similar list to learn about the geographic origins of the Jewish Russian immigrants; the current data adds a time dimension, in that it enables to observe the way the share of each region among the immigrants has changed over time.

are built up can hinder mass migration at the earlier stages. Polish and Lithuanian Jews were not only early starters, they kept forming more associations per-capita than Jews from the wealthier yet pogrom-stricken South-West and New-Russia throughout the entire period.

3.4 Descriptive Statistics

Currently, immigration flows from towns in 215 different districts were measured, almost all of them from the Pale of Settlement. These measures were aggregated at the cohort-district-year level and matched with the demographic data on each cohort from the 1897 census, as described above. The reported sample was then restricted to immigrants at ages 16-50 only, which will henceforth be referred to as the *sample age interval*. The *sample cohorts*, those that fell within the sample age interval in at least one of the *sample years* 1899-1913, were one year old some time during the period 1850-1898.⁴⁰

Table 2 reports the descriptive statistics of the main variables used in the estimation, and a few issues pertaining to it need to be dwell upon. Panel (A) reports the district-level variables: an average district had 44.7 migrations observed for each 1,000 individuals from among the sample cohorts throughout the sample period. Since the migrations that were observed are only a fraction of the total Jewish Russian migrations to the United States at that period, I use the yearly ratio of officially recorded to identified migrations to inflate the migration counts and arrive at an adjusted measure of migration.⁴¹ According to the adjusted measure, an average district had 147.1 migrations for each 1,000 individuals.

There are a handful of districts for which migration is probably over-estimated. In seven districts the estimated migration was more than half of the size of the respective cohorts, and in two of them migration had exceeded it. This reflects the most concrete danger to the correct identification of the size of migration flows within provinces: in almost all cases, these are the districts in which the province capital was located. For example, in the case of the city of Minsk, both the district and the province in which it was located were named after it. Immigrants arriving from the province of Minsk, but not from the district of Minsk, may have reported their last place of residence as simply "Minsk". The identification algorithm would have erroneously matched them to the city of Minsk and aggregate them into the migration flow coming from the district of Minsk. So there is a potential source of bias towards provincial capital districts.⁴²

⁴⁰ For example, the cohort that was one year old in 1890 is only observed migrating during the years 1905-1913 - starting at age 16 and up to the end of the sample period.

⁴¹ The ratio is calculated based on all migrations whose local origins are identified, including migrations outside the sample age interval and the sample cohorts.

⁴² Most, but not all provinces were named after their main cities. Another problematic singular case is the district of Grobin, in Courland province, that included the port city of Libava (Libau). This was the main Russian port for direct transatlantic embarkations. While most Jews could not use it for emigration, because legal embarkation from within the Russian Empire required a hard-to-acquire passport, a minority of them did. Some of them have reported Libau as their last place of residence, and thus the identified migration from this district appear grossly inflated.

There are several ways to tackle this problem. First, the estimation model allows for district random-effects, which hopefully captures, albeit imperfectly, the capital district bias. Another would be to try and estimate within the model the probability of an immigrant reporting the capital district instead of the actual district. Currently, I omit from the sample cases in which the rates of immigration from a district crosses a certain threshold.

However, the extent of the capital district bias should not be exaggerated. If the yet undocumented conjecture that more urbanized cohorts were more likely to migrate is correct, then there is a reason why migration from the capital districts should have been significantly greater than from the remaining districts. First, because the province capital was typically the province's largest urban center; and second, because the share of Jews within the district living in the district's urban center was larger in the capital districts.⁴³

Next on the descriptive statistics, an average district in the sample had almost two towns whose migration flows were identified, covering 59 percent of the total Jewish population of the district. To the extent that migrants tended to report the nearest large town if their hometown was small, the effective coverage rate must be higher than this figure.

Panel (b) reports statistics at the district-year level. The average yearly-district rate of migration, among the cohorts that are at the sample age interval during that year,⁴⁴ is 14 per thousand. It varies strongly across ages. Within the youngest five cohorts (ages 16-20) the average yearly-district rate of migration is almost 21 per thousand, three times greater than for the oldest five cohorts (ages 46-50). This intensity of migration among the young highlights the importance of accounting for the selection of stayers across time. The rates of migration among them was such that within a few years those who stayed behind would have remained a smaller and selected group of non-movers.

Panel (c) follows cohorts within each district, and again shows how significant the selection into migration among the young may have been. The average district-cohort had lost more than 13 percent of its 1897 population to U.S. migration during the sample period (while being at the sample age interval). But from among the 1884 cohort that came of migration age in 1899, the youngest that was followed through the entire sample period, almost one of four living in Russia in an average district in 1897 have relocated to the U.S. during the years 1899-1913. In a quarter of the districts this rate has exceeded 30 percent. The titular term *exodus* requires no further justification than these figures.

⁴³ For example, of 77 thousand Jews recorded in the capital district of the province of Vilna, 64 thousand were in the city of Vilna. As opposed to that, Vilna province's second largest city, Smorgon, was home to only 6,743 Jews of 28 thousand living in the district.

⁴⁴ Note that in each year this is a subset of the sample cohorts. For example, the 1890 cohort is not included in this calculation before it turns 16 in 1905.

4 Model

This section describes the estimation model I use. The purpose of this model is to describe the crucial features of the decision to migrate, in a way that could be mapped to available data and be used in estimation and capture the effects of income differences and business-cycles on individuals' decision to migrate. It enables linking the macro-level fluctuations in migration to micro-foundations. It does so by modeling migration as an optimal stopping time problem, as in Rust (1987), with finite horizon, an unobserved heterogeneity across individuals (types), and two unobserved time-varying individual state variables, representing individual cycles and linkage status.

4.1 A Prospective Migrant's Problem: Overview

There are two countries. The country of origin l^0 , and the country of destination l^1 . Individuals start their lives in the country of origin, and at any period t , individual i whose (deterministically advancing) age is a_{it} , expects to benefit from a periodical stream of utility flows $\{u_{it}, \dots, u_{i,t+\tau}, \dots, u_{iT_i}\}$ up to a finite known horizon T_i , in which he will reach the final age a_{\max} . Individuals living in the country of origin are forward looking, they form expectations about the future paths of utility in both countries based on current information. At each period they face a binary decision, *migrate* or *stay*, knowing that if they stay the option to migrate is still available at all future periods. However, migration is irreversible, once an individual migrated, there is no option to go back to the country of origin.⁴⁵ Thus, migration is modeled as a case of an optimal stopping time problem with a finite horizon.

A prospective migrant i who at the beginning of period t is found at the country of origin (i.e., $l_{i,t-1} = l^0$), then faces the problem of finding his optimal next location:

$$l_{it}^* = \arg \max_{l_{it} \in \{l^0, l^1\}} \mathbb{E}_t \left(\sum_{\tau=0}^{T_{it}} \beta^\tau u_{i,t+\tau} \right), \quad (3)$$

where β is the discount rate, and the dependency of utility on the location is suppressed in the notation.

To represent this problem in a Bellman equation, let s_{it} denote the vector of all state variables affecting current utility flows in both countries, except for age and location. Let $S_{it} \equiv \{\dots, s_{i,t-1}, s_{it}\}$ be the history, such that $(S_{it}, a_{it}, l_{i,t-1})$ fully describes the prospective migrant's information when facing the decision, as well as the determinants of his current utility flows in both countries.

The flow utility can then be written as a function $u_{it} = u(s_{it}, a_{it}, l_{i,t-1})$. Denote by the function $V(S_{it}, a_{it}, l_{i,t-1})$ the expected net present value of lifetime utility flows when starting period t at

⁴⁵ As discussed in section II, in the case of the Jewish migration the number of returning migrants was particularly small. Explaining why the rates of return migration should vary so much conditional on the gross rates of migration remains an economic challenge that has yet to be solved.

age a_{it} at location $l_{i,t-1}$, with history S_{it} and under optimal decisions. At each period, given S_{it} , a_{it} , and $l_{i,t-1} = l^0$, the expected net present value is

$$V(S_{it}, a_{it}, l_{i,t-1}) = \max_{l_{it} \in \{l^0, l^1\}} \{u(s_{it}, a_{it}, l_{it}) + \beta \mathbb{E}_t V(S_{i,t+1}, a_{it} + 1, l_{it})\}, \quad (4)$$

where the expectations are taken over $S_{i,t+1}$ with respect to S_{it} .⁴⁶ This represents period t 's value as the sum of period t 's utility flow and the discounted expected continuation value at period $t + 1$, under optimal decision making. The value function V is well defined and could be solved recursively by iterating backwards over ages starting from the last age a_{\max} when no continuation value exists, and in which the value function is simply

$$V(S_{it}, a_{\max}, l_{i,t-1}) = \max_{l_{it} \in \{l^0, l^1\}} u(S_{it}, a_{\max}, l_{it}). \quad (5)$$

Finally, the decision to migrate can be written as a policy function of the state variables:

$$\begin{aligned} l_{it}^* &= l(S_{it}, a_{\max}, l_{i,t-1}) \\ &= \arg \max_{l_{it} \in \{l^0, l^1\}} \{u(s_{it}, a_{it}, l_{it}) + \beta \mathbb{E}_t V(S_{i,t+1}, a_{it} + 1, l_{it})\}. \end{aligned} \quad (6)$$

The next section explains how the flow utility is determined, followed by a section describing how expectations regarding future flow utility are formed.

4.2 Flow Utility

The flow utility is a function of three arguments:

$$u_{it} = \check{u}(w_{it}, c_{it}, \varepsilon_{it}) \quad (7)$$

where w_{it} is individual's real income; c_{it} are the costs of migration, applying during the period in which migration takes place; and ε_{it} is a random term. All three components are themselves functions of the state variables, including age and location.⁴⁷

Specifically, I assume an additive function in which real income enters through the function v_i :

$$\check{u}_i(w_t, c_{it}, \varepsilon_{it}) = v_i(w_t) - c_{it} + \varepsilon_{it}. \quad (8)$$

Note that the costs of migration are expressed in terms of the utility measure, and not in terms of income. Below I describe how each argument depends on the primitive state variables.

⁴⁶ After migration, the case in which $l_{i,t-1} = l^1$, the function is trivially $V(S_{it}, a_{it}, l_{i,t-1}) = u(s_{it}, a_{it}, l_{it}) + \beta \mathbb{E}_t V(S_{i,t+1}, a_{it} + 1, l_{it})$, with $l_{it} = l^1$.

⁴⁷ So $\check{u}(w_{it}, c_{it}, \varepsilon_{it}) = u(S_{it}, a_{it}, l_{it})$; notations expressing this dependence are suppressed for clarity.

4.2.1 Real Income

Real income in each of the two countries follow independent growth processes with deterministic growth trends, where business-cycles are the deviations of actual real income from the trends. The log of income levels in each country, $\ln w_t^0$ and $\ln w_t^1$, are decomposed to the sums of trend levels and the deviations from the trends. Suppressing the country notations, the log of real income in each country is

$$\ln w_t = \ln \tilde{w}_t + d_t \quad (9)$$

where \tilde{w}_t denotes the trends and d_t the deviations. The marginal utility from income is denoted by α_w . On top of the general level of income in the country of destination $\ln w_t^1$, individual i 's utility from income in that country is augmented (or discounted) by an additional personal term η_i . In sum, the utility from income in the two countries can be written as:

$$v_i(w_t) = \alpha_w(\ln \tilde{w}_t + d_t) + \eta_i \mathbb{1}_{[l_{it}=l^1]}. \quad (10)$$

The individual fixed-effect η_i , multiplied by an indicator for the country of destination, stands for the prospective migrant's "type" on the scale of movers and stayers. The greater it is, the relatively better-off he is made by migration.⁴⁸ The model takes it as given that conditional on all observable characteristics, there remains an unobserved heterogeneity across individuals in their inherent propensity to migrate. It is agnostic as to the sources of this heterogeneity and as to the question whether migrants are positively or negatively selected.⁴⁹

4.2.2 Costs of Migration

The costs of migration depend on the state of the destination business-cycle, as well as on the age of the migrant:

$$c_{it} = [\gamma_0 - \gamma_1 d_t^1 + f_e(a_{it})] \mathbb{1}_{[l_{it}=l^1, l_{i,t-1}=l^0]}, \quad (11)$$

where the indicator is for migration in period t . The age-dependent component of the cost function is

$$f_e(a_{it}) = \kappa_e(a_{it} - a_{\min}) \sum_{\tau=0}^{T_{it}} \beta^\tau. \quad (12)$$

⁴⁸ There is no loss of generality in having an individual fixed-effect apply only in the country of destination. Given the additive specification, having two different effects, one for each country, is equivalent to a single effect in one country that could be thought of as the difference between them.

⁴⁹ This is a key question in the economics of migration. See Borjas (1987); for recent applications see McKenzie and Rapoport (2010), Abramitzky, Boustan, and Eriksson (2013, 2014), Gould and Moav (2014), and Spitzer and Zimran (2014).

where a_{\min} is the age in which a migrant starts his working life and is allowed to migrate, and $T_{it} = a_{\max} - a_{it}$ is the length of the remaining life at age a_{it} .

To be sure, the pecuniary costs are only a small part of the effective, broadly-defined, costs of migration. They are meant to capture anything from the difficulties of adjusting to a new environment and acquiring a new language; of moving to a new labor market in which not all previous labor experience may be relevant; or of the new constraints on the traditional ways of life, not the least on religious customs.⁵⁰ The destination business-cycles term captures the greater costs of migration during a recession, as discussed in section 2. This business-cycle effect is the primary effect driving the strong cyclicity of the flows of migration.⁵¹

The age-dependent component is meant to enable migration at older age to be more costly. Older prospective migrants may have greater losses in terms of non-transferable human and physical capital; they stand to lose location-specific labor experience, reputation, or connections with customers and suppliers;⁵² and they may be more likely to face complications in migrating as a household with children. The specification of f_e is derived from a simple Mincerian wage equation: The accumulated experience is $e_{it} = a_{it} - a_{\min}$, and the return to experience at each period is $\kappa_e e_{it}$. Upon migration, the migrant loses his entire previous experience, and starts accumulating new experience from scratch in the country of destination. The function $f_e(a_{it})$ is thus the present value of the lifetime loss of returns to the experience earned in the country of origin.⁵³

⁵⁰ Hatton and Williamson (1998, pp. 14-15) claimed that by the 1860s the decline in shipping costs had reached the point beyond which fluctuations in the prices had very little effect. Alroey (2008, Table 4C) calculated the total pecuniary costs of migration, including travel and expenses on the way to the port, on the order of \$100, where the fare of the ship ticket was under 40 percent of this total. This is a substantial sum in terms of Jews' income in the Pale—depending on the region typically more than half a year's wage (see Rubinow (1907, pp. 528-530))—but it could be covered within a few weeks of employment in the U.S. On the effects of changes in shipping fares, Keeling (2011) showed that the 1904 fare war induced many East-European migrants to embark through British and Scandinavian, rather than German ports. Deltas, Sicotte, and Tomczak (2008) claimed that cartelization in the transatlantic shipping market have had a significant effect in reducing emigration, but their evidence is not based on price data.

⁵¹ The specification could be extended, with significant computational costs, to include an analogous effect of origin business-cycles. However, it is probably the case that these effects are negligible compared to those coming from the business-cycles in the destination. Boustan (2007) showed that around that period the trends of migration from different European countries, such as Italy and the Austro-Hungarian Empire, followed almost precisely the same trend as that of Russian immigration, all corresponding strongly with U.S. income fluctuations. An evidence to the contrary comes from the case of the Swedish migration, many of whom were farmers susceptible to harvest shocks, that was also strongly correlated with domestic fluctuations; see discussion in Hatton and Williamson (1998, Ch. 4). Section 6 further discusses the importance of push factors in the case of Russian Jews.

⁵² This is particularly important given that over a third of the Jewish labor force in the Pale was employed in commerce, in which the intangible market-specific experience can not be used after migration.

⁵³ Chiswick (1991, 1992) showed that Jewish immigrants in New-York fared better in the labor market than other immigrants, and tended to climb up faster in the occupational ladder. Recent evidence by Abramitzky, Boustan, and Eriksson (2014) suggest that the seeming convergence of immigrants wages does not reflect convergence at the individual level, but is instead due to the decline in the quality of the cohorts over time and to selective return migration. Eckstein and Weiss (2004) estimated a model that explicitly captures lost and gradually reclaimed pre-migration experience.

4.2.3 Random Term

The random term ε_{it} is defined as follows:

$$\varepsilon_{it} = \zeta_{it} \mathbb{1}_{[l_{it}=l^0]} + \xi_{it}. \quad (13)$$

It comprises an individual country-specific I.I.D. transitory random term ξ_{it} , where the terms in the two countries are independent from one another, and an unobserved individual cycle term ζ_{it} , that only applies in the country of origin. This variable stands for the part of the individual variation over time that is not correlated with the macro-fluctuations, *and* has some persistence from one period to another. These could be temporary employment problems, a personal crisis, a recent migration of a close friend or relative, etc. It allows for some mobility across types over time. Conditional on observable characteristics and on the unobserved type, a prospective migrant can have persistent periods in which he is more or less prone to migration. The persistence of this cycle has bearing on the macro-patterns of migration: the greater it is, the more short-term persistent will be the stocks of prospective migrants from one year to another, meaning that a temporary shock to migration in one period will be compensated more strongly in the short-run by delayed migration.⁵⁴

4.2.4 Full Expression

After introducing all its components, the period t states vector could be defined formally as (suppressing the countries coupleness):

$$s_{it} \equiv (\tilde{w}_t, d_t, \eta_i, \zeta_{it}, \xi_{it}, l_{i,t-1}), \quad (14)$$

Finally, flow utility as a direct function of contemporary state variables can be written as (suppressing country-specific notations when possible):

$$\begin{aligned} u_{it} &= u(s_{it}, a_{it}, l_{it}) \\ &= v_i(w_t) + c_{it} + \varepsilon_{it} \\ &= \alpha_w (\ln \tilde{w}_t + d_t) + \eta_i \mathbb{1}_{[l_{it}=l^1]} \\ &\quad - \left[\gamma_0 - \gamma_1 d_t^1 + \kappa_e (a_{it} - a_{\min}) \sum_{\tau=0}^{T_{it}} \beta^\tau \right] \mathbb{1}_{[l_{it}=l^1, l_{i,t-1}=l^0]} \\ &\quad + \zeta_{it} \mathbb{1}_{[l_{it}=l^0]} + \xi_{it} \end{aligned} \quad (15)$$

Thus flow utility is a sum of utility from income, comprising trend, deviation from trend, and an

⁵⁴ Again, there is little loss in generality in assuming that the cycle applies in only one country. In practice, identifying individual cycle is difficult, and in the baseline estimation I shut it down.

individual fixed effect; of migration costs (if applicable), which are a function of the deviations from income trend and of age; and of a random term, composed of a transitory effect and an individual cycle effect.

Note that the business cycle variable d_t enters twice; once multiplied by the marginal utility from income α_w , and again multiplied by the coefficient of the variable costs of migration γ_1 . There is no perfect linearity, as the variable costs of migration enter the utility flow only during the period in which the individual migrates. Additionally, an exclusion restriction assumption makes the variable costs of migration depend only on d_t^1 , and not on d_t^0 . However, this assumption is strong and rather arbitrary, and the length of the two time series may not be sufficient to enable an identification of α_w based on the differences between them. As discussed above, the income trend $\ln \tilde{w}_t$ is a poor source of identification for the utility from income α_w , and thus in practice, despite the exclusion assumption and the fact that there is no perfect linearity, α_w and γ_1 are not separately identified. This is the same fundamental identification problem of separating the long-run from the short-term effects of income, discussed earlier in Section 2.2.1. In Section 6 I show how in principle the identification problem could be solved if sufficient data on panel or cross-sectional variation in real income is available, and what could be done in the absence of such data.

4.3 Transitions Between States

Given the history, age, and location $(S_{it}, a_{it}, l_{i,t-1})$, prospective migrants form expectations about the future state $s_{i,t+1}$, and make their decision on migration accordingly, as described by the policy function in equation 6. The following describes the transition of state variables from one period to another.

First, a few state variables have a trivial transition. The next location $l_{i,t}$ is determined by choice. Age advances trivially one period at a time, $a_{i,t+1} = a_{it} + 1$. And the unobserved type η_i is fixed.

The income trend \tilde{w}_t is assumed to have a fixed growth rate g (which may be different for the two countries), meaning that the log of income trend is linear in time. The deviations from trend d_t revolve about the trends in an AR(1) process, which may have different parameters in the two countries.⁵⁵ Thus, period t 's levels of trend incomes and the deviations from them are projected to be:

$$\ln \tilde{w}_{t+1} = \ln \tilde{w}_t + g, \tag{16}$$

$$d_{t+1} = \varphi d_t + \mu_{t+1}, \tag{17}$$

⁵⁵ This is a common way of describing business-cycles and growth paths, but it is by no means consensual (see the unit-root debate, following Nelson and Plosser (1982)). In particular, it implies that the effects of business-cycle shocks on future income is asymptotically zero. I chose the common AR(1) specification due to its simplicity and tractability, and due to the fact that it reflects beliefs that appear natural: business-cycles have some persistence, but decay over time.

where the random component of the AR(1) process is distributed normally and independently in each country, potentially with different variances:

$$\mu_t \sim \mathcal{N}(0, \sigma_\mu^2). \quad (18)$$

It follows that the transition distribution functions of the deviations from trends from period t to period $t + 1$ is:

$$\begin{aligned} F_d(d_{t+1}|S_t) &= F_d(d_{t+1}|d_t) \\ &= \Phi\left(\frac{d_{t+1} - \varphi d_t}{\sigma_\mu}\right) \end{aligned} \quad (19)$$

where Φ is the normal cumulative distribution function.

The transitory terms ξ_{it} are assumed to be distributed I.I.D. according to the Extreme Value distribution Type I, independently in the two countries and with the same variance:

$$\xi_{it} \sim \text{EVI}(0, \sigma_\xi). \quad (20)$$

Finally, the individual cycle term ζ_{it} has a simple binary distribution. It assumes one of two values and at each period there is a probability p to switch:

$$\zeta_{it} \in \{-\zeta, \zeta\}, \quad (21)$$

$$\zeta \geq 0, \quad (22)$$

$$\Pr(\zeta_{i,t+1} = -\zeta_{it}) = p. \quad (23)$$

The case $\zeta_{it} = -\zeta$ represents a temporary increase in the propensity of the individual to migrate, since he will benefit less from current flow if he stays in the country of origin. But as opposed to the transitory I.I.D. term ξ_{it}^0 , as long as $p < 0.5$ this condition is likely to persist into the next period.

In sum, according to the assumptions, all the information required to know the distribution of next period's state variables s_{t+1} is contained in current period's state variables s_t , and the previous history does not add predictive information over it. Thus states follow a Markovian process, $F_s(s_{t+1}|S_t) = F_s(s_{t+1}|s_t)$, that is defined by the following parameters:

PARAMETER	DESCRIPTION
g	Growth rates
σ_μ	Variances of B.C. AR(1) random terms
σ_ξ	Variance of transitory IID terms
p	Transition probabilities of individual cycle

The transition distribution function $F_s(s_{t+1}|s_t)$ is the one used while taking expectations in the Bellman equation 4 and the policy function 6.

4.4 The District Level

Up until now, the dynamic programming problem of an individual of type η_i was described. Much of the interesting phenomena of mass migration go beyond the individual level and depend on the way prospective migrants are aggregated. The current section gives an account of the features that pertain to the district-level: the variation of types *within* districts, and the variation of the distributions of types *across* districts.

Within each district j there is a different distribution of the unobserved types η_i . It is assumed to be normal, with the same variance in all districts, and centered around the district-specific mean η_j :

$$\eta_i \sim \mathcal{N}(\eta_j, \sigma_\eta^2). \tag{24}$$

This distribution is that of prospective migrants starting their lives at district j , but after the population of the district had been exposed to the risk of migration, this is no longer the distribution of the population actually living there. In particular, since movers are more likely to migrate, the mean of the distribution is expected to decline, increasing the share of stayers, and it might no longer be normal. This means that within each observables-based population bin, the greater the past probability to migrate, the lesser the probability of current migration.⁵⁶

This is the mechanism that drives the phenomenon of delayed migration: a shock to past migration changes the probabilities to observe migration in the following years, not only because the number of prospective migrants exposed to the risk is affected by the shock, but also due to the fact that its unobserved composition has changed. It is one of the challenges of the estimation procedure, described in the next section, to take this selection into account and properly discount current probabilities to migrate based on past migration.

The district-specific mean of the distribution is a function of the district characteristics x_j , which may be a scalar or a vector, and of a district random term ϵ_j . In particular, I assume a linear relation, with a normally distributed district random term:

$$\eta_j = \alpha_0 + \alpha_1 x_j + \epsilon_j, \tag{25}$$

$$\epsilon_j \sim \mathcal{N}(0, \sigma_\epsilon^2). \tag{26}$$

The district characteristics x_j are meant to capture differences in income across districts, and other

⁵⁶ That is, while controlling for the size of the network. This doesn't rule out that past migration had another effect in the opposite direction, through increasing the size of the network.

features that might affect the general tendency of the district’s population to migrate, such as the sort of occupations that are prevalent in it, literacy, etc. In Section 5, I show that it is also possible to estimate directly the vector of district fixed-effects.

4.5 Networks

The problem of optimal migration time applies from the moment in which a prospective migrant becomes linked, from which point he has an option to migrate. Prior to that, he must stay. This section describes the process by which individuals are becoming linked over time.

First, a link is an individual binary state, $k_{ijt} \in \{0, 1\}$, where $k_{ijt} = 1$ denotes that individual i in district j is linked at period t . Linkage is irreversible; once an individual becomes linked he knows that the option to migrate will be available through the rest of his life. This, and the fact that the linkage status is binary rather than continuous,⁵⁷ is a major simplification. But it comes with a great computational advantage, that the linkage process is removed from the dynamic programming problem. Individuals in the model need not form expectations about the future paths of their linkage status.

Another strong assumption made is that the probability to become linked at period t conditional on not having been linked before, the linkage rate K_{jt} , is the same for all unlinked prospective migrants, independently of their observed and unobserved characteristics.⁵⁸ Formally, the probability to be linked is defined as:

$$\Pr(k_{it} = 1) = \begin{cases} 1 & \text{if } k_{ij,t-1} = 1, \\ K_{jt} & \text{if } k_{ij,t-1} = 0. \end{cases} \quad (27)$$

The district linkage rate K_{jt} is a function of the current activity of the immigrants network at the country of destination. This is captured by the recent activity in the same district N_{jt}^d , and in the province to which the district j belongs, N_{jt}^p . Specifically, let $\{\dots, n_{j,t-1}^d, n_{jt}^d\}$ be measures of network activity at district j in periods $\{\dots, t-1, t\}$. Let $\{\dots, n_{j,t-1}^p, n_{jt}^p\}$ be measures of activity at the province, potentially outside district j .⁵⁹ Then the current network activity of district j at period t is defined as:

$$N_{jt}^d = \sum_{\tau=0}^{\infty} \delta^{\tau} n_{j,t-\tau}^d, \quad (28)$$

⁵⁷ That is, there are no different degrees of linkage strengths that make migration more or less costly. No linkage is equivalent to infinite migration costs.

⁵⁸ Assuming that this process is exogenous implies that prospective migrants can not actively look for links. Only once they are linked they respond differently to incentives.

⁵⁹ In practice, n_{jt}^d and n_{jt}^p are the new associations per capita founded at year t by immigrants coming from the district and from the province.

where δ is the rate of depreciation. The province-level activity N_{jt}^p is defined in a similar way.

The linkage probability then aggregates the recent network activity:

$$\begin{aligned} K_{jt} &= K(N_{jt}^d, N_{jt}^p) \\ &= \text{logit}(\lambda_0 + \lambda_d N_{jt}^d + \lambda_p N_{jt}^p). \end{aligned} \tag{29}$$

This characterization of the diffusion process is meant to capture the following: When the network of immigrants thickens, such as when immigrants from the district or from the province incorporate a new association, more links are formed. This means that the probability of a yet-unlinked prospective migrant to become linked at this period is greater. Similar network thickening that occurred in previous years may also increase this probability, but its effect depreciates over time, hence the depreciation factor δ . This specification is a simplified discrete-period analogue to the (Bass 1969) Diffusion Model.⁶⁰

Note that pioneers are allowed, in the sense that even if there is zero network activity, the probability to become linked is strictly positive. The greater is λ_0 , the more spontaneous links are being formed. Thus, in the absence of network formation there will still be an underlying drift of prospective migrants gradually being linked. This enables the estimation procedure to effectively test the assumption that networks matter; if the networks formation is nothing but noise, the parameter λ_0 would be estimated to be so large that within an instant of the beginning of the process all districts are linked. On the other hand this specification is restrictive in being deterministic, as it does not include a random term.

* * *

This concludes the description of the model. Table 1 summarize the variables and the parameters that govern it. In practice, some of the parameters are weakly identified, and thus the benchmark model that is used in the results section has several simplifications relative to the full-fledged model described above (see Section 6.2). The sensitivity of the main results to these simplifications will be tested in robustness checks.

⁶⁰ In the continuous-time Bass diffusion model, the momentary linkage (hazard) rate would be a linear function of the size of the effective network: $K_{jt} = \lambda_0 + \lambda_1 N_{jt}$, where the size of the network is replacing for the number of previous adopters in the Bass model. I use the logit transformation to bind the predicted probability in the range (0, 1). On estimating the Bass model in discrete time see Srinivasan and Mason (1986) and Boswijk and Franses (2005).

5 Estimation

I perform a Maximum Likelihood Estimation, in which the goal is to maximize the function $LL(\Omega|\theta)$, the log-likelihood to observe the data Ω conditional on the set of parameters θ . The next subsection is a concise overview of the way in which the log-likelihood function is being calculated. A complete mathematical description is on the estimation appendix. It is followed by a more detailed discussion of some of the issues in the estimation procedure and the results.

5.1 Main Steps in Estimation Procedure

Throughout this section I will use the following terms and notations in referring to time periods: The *initial stage*, denoted $T^{\text{init}} = \{t^{\text{init}}, \dots, 0\}$, spans the years in which migration took place and on which consistent macro-data are available, but no migration counts are available (FY 1885-1898). *Year-zero*, where $t = 0$, is the last year of the initial stage (FY 1898). Part of the estimation challenge is to predict the (end of) year-zero distribution of the population that has not yet migrated in terms of the unobserved variables. The *observed stage*, denoted $T^{\text{ob}} = \{1, \dots, t^{\text{ob}}\}$ is the era on which migration counts are available (FY 1899-1913). Finally, the entire era encompassing both stages is denoted $T = \{t^{\text{init}}, \dots, 0, 1, \dots, t^{\text{ob}}\}$. Further notations are $y \in Y$, denoting a year-of-birth cohort; $j \in J$ denoting districts; and $l \in \{l^0, l^1\}$ denoting the two countries.

The observed data Ω are

$$\Omega = (M, M^-, n, x, \tilde{w}, d), \quad (30)$$

defined as follows:

M	$= \{M_{yjt}\}_{y \in Y, j \in J, t \in T^{\text{ob}}}$	Cohort-district-yearly migration counts
M^-	$= \{M_{yj}^-\}_{y \in Y, j \in J}$	Cohort-district non-migration counts
n	$= \{n_{jt}^d, n_{jt}^p\}_{j \in J, t \in T}$	District-yearly network indicators
x	$= \{x_j\}_{j \in J}$	District characteristics
\tilde{w}	$= \{\tilde{w}_t^l\}_{l \in \{l^0, l^1\}, t \in T}$	Country-yearly income trends
d	$= \{d_t^l\}_{l \in \{l^0, l^1\}, t \in T}$	Country-yearly deviations from trends

The migration counts M are available only through the observed stage. The non-migration M^- is the residual from subtracting total observed migration from the size of the cohort observed in the 1897 census. The time-varying variables n , \tilde{w} , and d , are observed through the entire era T .

The estimation parameters vector θ is

$$\theta = (\alpha, \beta, \gamma, \delta, \kappa, \lambda\zeta, \sigma_\xi, \sigma_\eta, \sigma_\epsilon, p), \quad (31)$$

and its parameters are listed on Table 1, Panel B.

The estimation procedure follows through the following steps. For each set of parameters θ :

- (a) *Dynamic programming problem*: Solve the dynamic programming problem and compute the function

$$m(\eta, \zeta, a, \tilde{w}, d | k = 1, l = l^0),$$

the probability to migrate as a function of all state variables except for the transitory EVI term ξ , conditional on being linked and on not having migrated before.⁶¹ Plug in the actual values of the business cycle variables d , and rearrange this function, such that the age a and trend income \tilde{w} will be replaced with the cohort y and the period t , to get

$$m_{yt}(\eta, \zeta | k = 1, l = l^0),$$

the cohort-yearly probability to migrate as a function of type and individual cycle, conditional on being linked and present at the country of origin at the beginning of the period.

- (b) *Initial conditions*: Solve for the year-zero conditional distribution of unobservables, $F_{yj0}(\eta, \zeta, k | \epsilon, l = l^0)$. This is the probability that an individual of cohort y in district j at period t , will be of type η under individual cycle state ζ , with linkage status k , conditional on the district random-effect ϵ and on not having migrated prior to year-zero. Do that by following steps i-iv:

- (i) Draw Q randomly simulated paths $Z^{\text{init}} = \{z_1^{\text{init}}, \dots, z_Q^{\text{init}}\}$, where each path $z_q^{\text{init}} = \{\zeta_{qt^{\text{init}}}, \dots, \zeta_{q0}\}$ is a possible history of the serially correlated variable ζ over the entire initial stage, while using the transition probability p to generate draws. Use the solution to the dynamic programming problem to solve for the probability distribution function $F_{y0}(\zeta, k | \eta, t^k, z_q^{\text{init}})$, defined as

$$F_{y0}(\zeta, 1 | \eta, t^k, z_q^{\text{init}}) = \begin{cases} 0 & \text{if } \zeta_{q0} \neq \zeta \text{ or } t^k > 0, \\ \prod_{t=t^k}^0 [1 - m_{yt}(\eta, \zeta_{qt} | k = 1, l = l^0)] & \text{if } \zeta_{q0} = \zeta \text{ and } t^k \leq 0, \end{cases} \quad (32)$$

$$F_{y0}(\zeta, 0 | \eta, t^k, z_q^{\text{init}}) = \begin{cases} 0 & \text{if } \zeta_{q0} \neq \zeta \text{ or } t^k \leq 0, \\ 1 & \text{if } \zeta_{q0} = \zeta \text{ and } t^k > 0. \end{cases} \quad (33)$$

This is the probability of individual of cohort y , type η , who was linked at year t^k , and had gone under the individual cycle path z , to be present at the county of origin at year-zero, under individual cycle state ζ , and with linkage status k .⁶²

⁶¹ This section uses a series of migration probabilities functions, all denoted with m , differing from each other by the arguments and the conditioning. The dependence on the set of parameters θ and the subscripts of the variables are suppressed for clarity.

⁶² Note that the linkage status k follows directly from t^k , and that ζ is just the last element of the path z .

(ii) Integrate the function F_{y0} over the paths z and get the probability distribution function

$$F_{y0}(\zeta, k|\eta, t^k) = \frac{1}{Q} \sum_{q=1}^Q F_{y0}(\zeta, k|\eta, t^k, z_q^{\text{init}}). \quad (34)$$

Note that when $t^k > 0$ (i.e., the individual is not yet linked in period zero), the probabilities are simply $F_{y0}(\zeta, k|\eta, t^k) = 0.5 \cdot \mathbb{1}_{[k=0]}$. This is the “virgin” distribution, yet unaffected by attrition caused by past migration.⁶³

(iii) Given the linkage parameters λ , plug the network measures n in the linkage functions in equations 28 and 29 to find the district-yearly linkage rates K . Compute $P_j^k(t)$, the probability to become linked in district j at year t :

$$P_j^k(t) = K_{jt} \prod_{\tau=t^{\text{init}}}^{t-1} (1 - K_{j\tau}), \quad (35)$$

and denote the probability to not have been linked by year zero by

$$P_j^{-k}(0) = \prod_{\tau=t^{\text{init}}}^0 (1 - K_{j\tau}). \quad (36)$$

Integrate $F_{y0}(\zeta, k|\eta, t^k)$ over t^k with respect to the probabilities $P_j^k(t^k)$, and get for each cohort and district the year-zero probability distribution

$$F_{yj0}(\zeta, k|\eta) = \begin{cases} \sum_{t^k=t^{\text{init}}}^0 F_{y0}(\zeta, k|\eta, t^k) P_j^k(t^k) & \text{for } k = 1, \\ 0.5 P_j^{-k}(0) & \text{for } k = 0. \end{cases} \quad (37)$$

(iv) Use $F_{yj0}(\zeta, k|\eta)$ and $F(\eta|\eta_j)$, the prior probability to be of type η conditional on the district mean type η_j , to solve for the year-zero distribution of unobservables among individuals who have not migrated during the initial stage, conditional on the district mean type:

$$F_{yj0}(\eta, \zeta, k|\eta_j, l = l^0) = F_{yj0}(\zeta, k|\eta) F(\eta|\eta_j). \quad (38)$$

(c) *Conditional probabilities to observe migration:* Solve for $m_{yjt}(\epsilon|l_0 = l^0)$, the probability of individual of cohort y in district j to be observed migrating at period t , when the district random-effect is ϵ , conditional on being present through year-zero. To do that, follow steps i-iii:

(i) Draw Q randomly simulated paths $Z^{\text{ob}} = \{z_1^{\text{ob}}, \dots, z_Q^{\text{ob}}\}$ where each path $z_q^{\text{ob}} = \{\zeta_{q,1}, \dots, \zeta_{q,t^{\text{ob}}}\}$ is a possible history of the serially correlated variable ζ over the entire observed stage,

⁶³ The probabilities are 0.5 because the transition probabilities of the process ζ are symmetric.

Use the solution to the dynamic programming problem and integrate over the paths to get

$$m_{yt} \left(\eta, \zeta_0, t^k | l_0 = l^0 \right),$$

the probability to migrate as a function of type η , year-zero individual cycle state ζ_0 , and linkage year t^k , conditional on being present through year-zero.

- (ii) Use the year-zero distribution F_{yj0} that was found on (b) and the linkage processes K , to integrate $m_{yt} \left(\eta, \zeta_0, t^k | l_0 = l^0 \right)$ over the year-zero individual cycle status ζ_0 and the linkage years t^k , and get

$$m_{yjt} \left(\eta | l_0 = l^0 \right),$$

the probability to be observed migrating as a function of the type η and conditional on being present through year-zero.

- (iii) Integrate $m_{yjt} \left(\eta | l_0 = l^0 \right)$ over η with respect to the district random-effect ϵ and the district characteristics x , using the probability distribution F_{yj0} that was found on (b). Get the cohort-district-year probabilities to migrate as a function of the district random-effect

$$m_{yjt} \left(\epsilon | l_0 = l^0 \right),$$

and back out from it the probabilities to not migrate,

$$m_{yj}^- \left(\epsilon | l_0 = l^0 \right).$$

- (d) *Likelihoods*: Compute the log-likelihood $LL(M, M^-, N, x, \tilde{w}, d | \theta)$.

- (i) Using the probabilities to migrate and to not-migrate found above, the likelihood of observing the observed pattern of migration from cohort y at district j , given a district random-effect ϵ , is (suppressing the conditioning on $l_0 = l^0$ and the observed variables that were used earlier on):

$$L_{yj} \left(M_{yj}, M_{yj}^-, \epsilon; \theta \right) = \prod_1^T \left(m_{yjt} \left(\epsilon \right)^{M_{yjt}} \right) m_{yj}^- \left(\epsilon \right)^{M_{yj}^-}. \quad (39)$$

- (ii) Integrating the conditional cohort-district likelihoods over the random-effect ϵ , assumed to be distributed normally as discussed above, the district j likelihood is:

$$L_j \left(M_j, M_j^-; \theta \right) = \int_{\epsilon} \prod_{y \in Y} \left(L_{yj} \left(M_{yj}, M_{yj}^-, \epsilon; \theta \right) \right) f_{\epsilon}(\epsilon) d\epsilon \quad (40)$$

where $f_{\epsilon}(\epsilon)$ is the density of the normal distribution with variance σ_{ϵ} .

(41)

(iii) Finally, taking logs and summarizing over district, we get the desired log-likelihood function:

$$LL(M, M^-, N, x, \tilde{w}, d|\theta) = \sum_{j \in J} \ln L_j(M_j, M_j^-; \theta) \quad (42)$$

6 Results

6.1 Determinants of Migration: Reduced Form Regressions

The main patterns of correlations between migration and its potential determinants can be observed in Table 3. The table reports OLS regressions at the district-year level, similar to traditional push-pull analysis of migration, predicting the rate of migration among the cohorts aged 16–50 at each year, according to the following specification:

$$m_{jt} = \alpha + \beta x_j + \gamma w_t + \delta z_{kt} + \epsilon_{jt}, \quad (43)$$

where x_j are district time-invariant characteristics, w_t are time-series variables (income or year fixed-effects), and z_{kt} are the yield shocks in year t at province k , to which district j belongs. In columns 1-4 the dependent variable m_{jt} is yearly migration per thousand, and in columns 5-8 it is the log of yearly migration per thousand.

The top four rows report the coefficients of district characteristics. In the absence of a direct district-level measure of income, a reasonable variable representing the local conditions that affect the propensity to migrate is the ratio of Jewish workers in commerce to Jewish workers in manufacturing, appearing at the top row.⁶⁴ According to columns 1 and 5, manufacturing districts indeed sent more migrants.⁶⁵ Another potential cause for an association between this variable and the rate of migration is that in urban centers there were relatively more manufacturing workers. The negative correlation could be a result of greater inclination of urban cohorts to migrate, independently of the standards of living. Furthermore, it is possible that there was selection into migration based on occupation, if artisans and laborers were more likely to benefit from migration compared to workers in trade and commerce.⁶⁶ Finally, this variable might reflect the geographic dispersion of migrants, as the northwestern districts had both higher rates of manufacturing workers and of U.S. migration.⁶⁷ Indeed, the fact that the coefficient becomes much smaller and statistically insignificant when province fixed-effects are added (columns 3 and 7), suggests that this correlation was largely driven by regional patterns.

Capital districts have greater counts of migration. This could be attributed either to greater propensity to migrate in urban centers, or to migrants reporting their province’s main city rather than their own town of origin—a potential source of measurement error. As expected, districts with more comprehensive coverage in the migration data also had greater migration counts. The positive correlation between these two variables and migration, which remains even after controlling

⁶⁴ As mentioned above, it was observed that the material conditions of the Jews were worse in regions in which more of them found their way out of the traditional commerce niche into manufacturing jobs.

⁶⁵ The coefficient of -0.21 in column 5 means that an increase of one standard deviation in the commerce/manufacturing ratio is associated with 19 percent fewer migrants.

⁶⁶ This hypothesis was critically discussed in Perlmann (2000).

⁶⁷ See yanayspitzer.net/2012/09/30/jewish-occupations-in-the-pale-of-settlement.

for province fixed-effects, is thus likely to be driven to a large extent by systematic time invariant district-specific measurement errors. This highlights the importance of controlling for district fixed-effects in the estimation. The cumulative rate of associations per capita founded through the period prior to the first year of observed Ellis Island migration (1861–1899), stands as a measure of prior penetration of migration networks. As expected, it is strongly correlated with migration during the observed period, although this correlation decreases after controlling for province fixed-effects. The lower within-province correlation is consistent with the notion that the networks’ spheres of influence crossed district boundaries.

As is almost invariably the case in the literature (and as shown for the case of Russian Jews in Boustan (2007)), American business cycles are strongly positively correlated with migration. The coefficient of 9.78 (column 5) implies that a 1 percent greater US income is associated with 10.3 percent more migrants. However, Russian income (added to the regressions in columns 2 and 6), is not correlated with migration in the expected direction. This does not appear to be a result of correlation between the US and Russian cycles, as the coefficient on US income remains virtually unchanged. This seemingly perverse pattern may be a result of the short length of the series. Nevertheless, it is often the case that the push effects on migration are rather elusive and dominated by the pull effects.⁶⁸

Further evidence on the effect of income push factors is given by the yields data. The measure of area-weighted averages of the shocks to yield-seed ratio of five primary crops is added to the regressions in columns 4 and 8, while controlling for district and year fixed-effects. As explained above, this measure is meant to reflect the exogenous weather-driven component of the the province-year shocks to agricultural income, which comprises the lion’s share of total income. The correlation between the yield-seed weighted residuals and migration is negative as expected, but it is weak and statistically insignificant (column 4) or virtually zero (in the log regression, column 8). It cannot be ruled out yet that this lack of correlation is merely a result of weak statistical power; the coefficient of the yields shock in column 8 can be interpreted in terms of elasticity, and its standard error is 0.092, such that the regression will fail to identify as statistically significant a correctly estimated yield elasticity of 0.18 or less. However, these levels are two orders of magnitude lower than the implied elasticity of US income, and although it is unknown how a yield shock translated into an effective income shock for the average Russian Jew, it is hard to imagine that it was attenuated more than fifty-fold.⁶⁹

To see this graphically, Figure 6 depicts nonparametric residuals regression of migration on yield shocks at the province-year level. The migration measures here are residuals from a regression of province-year log migration on province and year fixed-effects. Nothing gives the impression that there is any substantial negative relation between the two variables. The average migration

⁶⁸ See discussion in Hatton and Williamson (1998).

⁶⁹ If a 10 percent yield shock was transmitted into a 0.2 percent income shock only (a fifty-fold attenuation), and the push income elasticity was identical to the estimated pull income elasticity (9.7), then a correctly estimated coefficient of yield shock (0.194) would be just within statistical significance of 5 percent.

residual at the bottom quartile of yield shocks is in fact slightly lower than at the top quartile.⁷⁰ In short, at least in this case, the pull factors are dominant in determining fluctuations in migration, whereas the push factors are indiscernible. Failing to associate variations in income in the origin with variations in rates of migration does not prove that such causal link does not exist, but it does provide motivation for a model in which temporal variations in destination income are given precedence, such as by affecting the costs of migration more than the counterpart variations in origin income do. As discussed earlier, attempting to separately identify these long term effects of income difference from the short term effects of income fluctuations is subject to acute problems of identification within the traditional OLS framework.

6.2 Estimated Parameters: The Benchmark Model

The specification of the benchmark model has several simplifications compared to the complete model described in Section 4. First, the individual cycle is not yet introduced, meaning that individuals are assumed to have had no fluctuations in their underlying demand for migration.⁷¹ In experimentation, this process proved to be unidentifiable, and hence later robustness checks will test whether the main results are robust to assuming such process with arbitrary parameters.

Second, the mean utility intercept α_0 and the fixed costs of migration γ_0 are not separately identifiable. This means that the estimation cannot distinguish between a recurrent greater utility flow in the country of origin and a one-off greater cost of migration, and thus the effective costs of migration and the difference in average utility between the country of origin and the country of destination are not nailed down. Thus, the benchmark specification normalizes the fixed costs to $\gamma_0 = 0$. Moreover, since this specification enables district fixed-effects, such that each district's mean flow utility is estimated separately, the Panel-wide mean utility α_0 makes a redundant degree of freedom. Hence, it is set to $\alpha_0 = 0$.

The most undesirable, yet unavoidable, limitation of the full-fledged model is that in practice, the utility from income α_w is not identified separately from the effect of US business cycles shocks on the costs of migration, γ_1 . In fact, when estimating the two separately, the estimate of α_w is negative. This is due to the fact that based on the exclusion assumption, the only independent source of identification for this parameter is the series of Russian business cycles shocks, which as shown above, is correlated with migration in the “wrong” sign also in the reduced form regressions.

Intuitively, the problem is the following: a business cycle shock affects migration through two channels. It changes income in the current period and the expectation for future income in the future periods; it also changes the one-time costs of migration. The two effects cannot be separately identified, but they could be jointly identified in a reduced form specification. Since only the

⁷⁰ In a study on Swedish emigration, Bohlin and Eurenus (2010) found statistically significant yet “practically unimportant” effect of crop per hectare on emigration.

⁷¹ In practice, this means calibrating $\zeta = 0$.

first effect is related to the long-run effects of income differences on migration, a reduced form specification that bundles the two does not enable to estimate directly the long-run income effects. However, the reduced form specification does enable to estimate the short run effect of a temporary income shock, coming jointly through both channels.

Finally, given the questionable correlation between Russian business cycles d_t^0 and migration, this variable is set to zero, equivalent to assuming that Russian income was on trend at each period, or that the push effect of Russian income was zero. I experiment with alternative specifications as robustness checks to verify that the analysis is not affected by this strong assumption.

The benchmark specification of the flow utility is thus the following (compare to the full model in equation 15):

$$\begin{aligned}
u_{it} &= u(s_{it}, a_{it}, l_{it}) \\
&= [\kappa_t t + \eta_i] \mathbb{1}_{[l_{it}=l^1]} \\
&\quad - \left[\gamma_0 - \gamma_1 d_t^1 + \kappa_e (a_{it} - a_{\min}) \sum_{\tau=0}^{T_{it}} \beta^\tau \right] \mathbb{1}_{[l_{it}=l^1, l_{i,t-1}=l^0]} \\
&\quad + \xi_{it}
\end{aligned} \tag{44}$$

where κ_t replaces α_w and stands as a time trend coefficient.⁷²

Table 4 reports the parameters estimates of the benchmark model. As far as the parameters of the flow utility term, the outstanding thing to note is that the return to experience is estimated to be negative, a non-intuitive result. There are four channels through which age affects the probability to migrate. First, older migrants have shorter horizon, and thus a lower benefit from migration. Second, older prospective migrants are more likely to be stayers, as the movers are more likely to migrate early. Third, the time trend also reflect forces that make the return to migration vary over time, and thus by age. The loss of experience gained in the country of origin is the fourth channel. One way to interpret this result is that the effect of the experience term is hard to separately identify from any of the other three channels. In any case, the negative coefficient does not imply that the probability to migrate increases with age—other channels more than offset it.

In the diffusion process, the depreciation coefficient is statistically significant, but is close to zero. This means that a given expansion of the networks has an effect that brings about a one-off shift to the level of linkages, but does not accelerate the rates of linkage in the long run. Interestingly, the effect of network formation at the district level is dwarfed by that of the province level. This is likely to be a reflection of the noisiness of the associations measure at the district-year level. In most

⁷² Enabling a time trend in the country of destination is econometrically equivalent to leaving a utility from income term that is applicable only to the income trend (i.e., $\alpha_w \ln \tilde{w}_t$, not multiplied by the indicator). This term could be interpreted now either as a correction to the estimated income trends, or as simply a bundling of the differential income trends plus anything else that changes consistently differently over time in the two countries.

cases this is zero, and it infrequently receives a positive value. The province-level network formation averages over several districts, such that it is a more smooth measure, and by not suffering from very large artificial variation it could better represent the actual effects of networks.

6.3 Diffusion of Networks

The current section will describe several insights that follow from the estimation of the diffusion process within the benchmark model. As described in section 4.5, the option to migrate disseminates within districts and provinces; the greater is the prior exposure of the district and of other districts within the same province to migration, the more likely are individuals to be linked to a person who had already migrated, and thus to receive an “option to migrate”. Once such option is received, each individual solves his personal dynamic programming problem, and given his type and the state of the economies makes a decision each year whether to migrate or to wait.

The main variable of interest is the district-year linkage rate. This is the share of individuals originated in each district that have already received an option to migrate by a given year (some of them may have already migrated). By the assumptions of the model, the probability to be linked is the same for all individuals within a given district at a given year, independently of observables and unobservables. When the linkage rate approaches 1, the district is said to be saturated, in the sense that almost all individuals have already received migration options and very few new links can be formed.⁷³ The linkage process begins back in 1861, the first year in which a foundation of an immigrant association was identified.⁷⁴ Henceforth, subsequent foundations of associations bring about further linkages of previously unlinked individuals in the district and the province to which the associations pertain.

6.3.1 Patterns of Linkage

Figure 7 shows the estimated distribution of linkage rates across districts over the period 1861–1913. The primary patterns it reveals are the following: (a) districts start out unlinked, and can remain so over a long period; (b) districts can approach saturation, and by 1913 all of them do; (c) there is a range of partial linkage rates (i.e., the distribution of linkage is not binary—with districts either saturated or unlinked—instead, the case of intermediary levels of linkage is rather common); (d) during the observed period (1899–1913) most, but not all, districts are saturated, and the remaining districts are gradually becoming saturated over the period.

None of these patterns are mechanically produced by the assumptions of the model; they are informative non-obvious findings that characterize the evolution of migration over time and space. Put

⁷³ The specification does not allow virtual full saturation, i.e., linkage = 1, but districts can approach this rate.

⁷⁴ Some of the cohorts of interest were not yet born in 1861, yet in the model they could have benefited from pre-birth linkage. This process could be thought of as a household-level linkage, where individuals born to a linked household are already linked at birth.

together, they indicate that the diffusion process has a bite in the data. If migration probabilities were independent of prior network formation, then estimates that predict immediate network saturation would fit the data best. If, on the other hand, denser networks caused greater probabilities to migrate, but as in traditional models estimating migration this effect was linear (that is, not reaching saturation at any point), then the parameter values fitting the data best would not have enabled most districts to approach saturation. In other words, the traditional view of the networks effect could be nested within the estimation model used here, but the estimated parameters rule against it.

This diffusion model of networks provides different predictions than the traditional model. It predicts that after a while, the rate of migration would reach a plateau even in the absence of any decline in the incentives to migrate.⁷⁵ Accordingly, districts that have innate tendencies to produce high levels of migration may not fulfill any of this potential for a long period unless linked. There is, indeed, a drift of spontaneously generated links: in the absence of indicator of previous network formation, the rate of linkage is still predicted to grow. This drift, however, acts slowly. A hypothetical district that did not have any sign of associations related to it would have had a linkage rate of 48.7 percent by 1913. At this extreme case, after more than half a century, more than half of this district's potential for migration would still be unlinked.⁷⁶

6.3.2 Identification of the Diffusion Process

Since there is no direct migration data on the period prior to 1899, the patterns shown on it in Figure 7 are effectively out-of-sample extrapolations. To understand better what in the data leads the estimation procedure to identify the parameters of the diffusion process, it is useful to consider the the period 1899–1913. This is the period for which both the data on network histories (the associations data) and direct data on migration exist. At the beginning of this period, most districts are all but saturated, yet quite a few of them have yet to make more progress: One quarter of the districts had under 80.2 percent linkage, and 11.2 percent of the districts were below half way to saturation at this year.

It turns out that districts that were more weakly linked as of 1899, and hence becoming fully saturated during, and not before, the period of observed migration FY 1900–1914, have experienced far greater increase in the rates of migration during the observed period. Figure 8 shows this

⁷⁵ Hatton and Williamson (1998) propose several explanations for the typical inverse U-shape of the rate of migration during the Age of Mass Migration. These explanations hinge on various sources that reduce the incentives to migrate after the process of mass migration had matured. Network saturation could be added as an internal dynamics checking the growth of migration, independently of changes to the underlying incentives to migrate.

⁷⁶ The network formation is taken in the estimation model as an exogenous process, but clearly in practice it is not: Sooner or later, even migrants from such slowly linking districts would start their own associations. This case is extreme in the sense that it assumes that it is possible that the little linkage that is formed will not feed back to increase the speed of future network formation. Nevertheless, some districts in the data are estimated to have no more than 50 percent linkage as late as the turn of the century, four decades since pioneer migrants started leaving the Pale.

differential acceleration in migration, using a nonparametric regression of increase in migration on initial linkage. The horizontal axis represents the districts estimated rate of linkage as of 1899. The vertical axis shows the change in the rate of migration between the first seven years of the observed period (FY 1900–1906) and the later eight years (FY 1907–1914). The downward sloping curve means that on average, saturated districts experienced very little increase in migration, whereas the latecomers had been rapidly growing (more than 50 percent increase at the bottom of the 1899 linkage ladder). This pattern is possibly a source of identification for the parameters of the diffusion process. It is consistent with a reality in which districts with more advanced network formation are saturated, whereas lagging districts are still becoming linked, which is exactly the pattern predicted in Figure 7 for the last 15 years of the sample.

6.3.3 Geographic Patterns

More information on the geography of the diffusion process could be learned from Figure 9. The two plots depict yearly cumulative distribution functions of estimated linkage across districts, as in Figure 7, separately for the two regions of Poland and New Russia. The former was adjacent to the German border, and some of its districts were the pioneers of the Jewish migration, becoming linked and saturated during the 1860s and 1870s. Other districts within Poland, mainly those in the southern parts of Poland bordering Austria, were lagging behind. By the late 1890s, most Polish districts have become saturated.

In contrast, New Russia was further away from the earliest geographic origins of migration. The first signs of network formation does not appear before the late 1880s, and only by the 1910s did the region as a whole reach saturation. It seems that the linkage process in New Russia was lagging a decade or two behind that of Poland.

As discussed earlier, Kuznets (1975) and others pointed out that the fact that the northwest seemed to be over-represented among the Jewish migrants was a result of the poorer standards of living in this region compared to New Russia. But if New Russia was simply late to be linked, it could still be that its underlying demand for migration was equally high. This exemplifies the potential importance of controlling for the diffusion of networks when estimating the demand for migration: Variations in distance from the early sources of migration, and thus in the timing of linkage, could be mistakenly interpreted as variations in levels of propensity to migrate. This is the rationale behind the diffusionist view of the European pattern of transatlantic mass migration—the nations of southern and eastern Europe may have been latecomers simply because they were further away from the earlier European origins of migration, and not necessarily because their underlying demand for migration was low early on and rising rapidly around the turn of the century.

6.4 District mean types

The variation in the potential for the rates of migration—as opposed to the realization of the rates of migration, which may be inhibited by low linkage rates—is accounted for by district fixed-effects. As described in section 4.2.1, the average type in district j is η_j . The higher the type, the greater is the underlying demand for migration. This is the only time-invariant district-specific variable, and it functions as a “mean-utility” capturing anything that would shift the level of demand for migration in the district, net of the linkage process and the business cycles fluctuations. As explained in section 5, the estimation procedure produces a vector of estimates of these mean types. Figure 11 shows the distribution of these estimated district effects, which appears to approximate a normal distribution.⁷⁷

6.4.1 Long Run Income Effect on Migration: Prospects for Identification

In principle, if there is sufficient data on the level of income at each district, the correlation between the mean type and the level of income could be a source of identification of the long term effects of income on the rates of migration. For example, to know how a 10 percent increase in the level of real income at the district of origin reduces migration rates, the econometrician would need to identify the decrease in the district mean type caused by this increase in income, and then simulate the difference in the rates of migration resulting from the shift in the types distribution within the district. Unfortunately, the data currently available is insufficient to perform this exercise, and the remaining discussion will be mainly dedicated to the effects of short-term income shocks. However, the model used here provides a feasible method to identify the long-run effects of income on migration, while controlling for both the diffusion of networks and short-term income fluctuations.

One caveat to this is that the determinants of the network formation are not specified in the model, and are taken to be exogenous in the estimation. This means that as long as the districts are far from saturation, the model does not account for the feedback from greater migration to faster network formation, and thus would underestimate the effects of income differences. But once the districts enter a regime of full saturation, as is shown to be the case by 1914, the feedback from migration back to networks becomes negligible. Thus, identification of the long-run effects of income on migration is feasible for the case of saturated networks. Nevertheless, the estimation could be based on pre-saturation data.

⁷⁷ The mean types need not be distributed around zero. Since the types distribution is not separately identifiable from either the fixed costs of migration γ_0 and the utility intercept α , they were both normalized to zero such that the location of the types distribution is determined.

6.4.2 District Mean Type and Migration

To understand the role played by the district mean types and their interaction with network formation, consider Figure 11. Both plots present a scatter of the log of the average rate of migration against the estimated district mean type in the periods FY 1900–1906 (top) and FY 1907–1914 (bottom). That both scatters form a very tight upward-sloping curve is by construction: the mean types reflect all time-invariant differences between districts; districts with average high migration would generally be assigned with higher mean type, and vice versa. Correspondingly, the R-squared are around 0.9, meaning that almost all of the variation in the average rates of migration is explained away by the fixed-effects. The very wide variation of rates of migration is a pattern that was found repeatedly in migration studies—there is a very large unexplained regional variation across districts, and economic historians have always found it hard to explain by observed variable (Baines 1995).

The remaining variation around the curve is due to two causes: first, random differences in the realizations of average migration in the two periods; and second, the degree of network formation. If there was full saturation in both periods, the variation around the curves would have been the same in both plots. However, the curve at the bottom plot, referring to the later part of the period, is more tight than the top plot. The R-squared increases from 0.877 to 0.929, and the root of the mean squares decreases by almost a half. The mean type represents the potential for migration that could be achieved at saturation. As shown above, during the earlier period many districts are below saturation and their migration is therefore below their maximum potential. During the later period, almost all districts achieve saturation and thus converge to their potential. This is reflected in the plot by a tightening of the migration–types curve.

The variation around the curve appears to be of second order compared to the variation across the curve, even in the earlier top plot. But this is mainly due to the fact that the observed data is only from the last 15 years of the pre-WWI Jewish migration, the final moments of the Pale-wide linkage process. By the turn of the century, not all districts were saturated, but almost all were mostly linked. Extrapolating backwards to the 1880s and before, when only a handful of districts had been linked, the model would predict a far greater variation in the rates of migration given the level of mean type.

6.5 The Individual Decision on Migration

6.5.1 Types and Probabilities

A key component in the model is the unobserved variation of individuals’ types. Recall that types are defined according to the individual flow utility at the country of destination: “movers” receive a larger utility flow, denoted η_{ij} , compared to “stayers”, at each period they spend in the country of destination. The average type in district j is η_j , and within each district individuals’ types η_{ij}

are normally distributed around the mean: $\eta_{ij} \sim \mathcal{N}(\eta_j, \sigma_\eta^2)$. The fixed-effects specification enables to estimate the vector of the district fixed-effects, as well as the within-district variation σ_η . Using these estimates, I simulate a Pale-wide distribution of types.

In Figure 12, the horizontal axis represents the Pale-wide types percentiles (e.g., the range 90-100 represents the 10 percent of individuals most prone to migration). The probabilities to migrate are matched to each type from the solution to the dynamic programming problem, assuming that these individuals belong to the cohort that is 20 years old at FY 1900, and that they are present and linked at that year. The right curve represents the probability of each type-percentile to migrate at age 20 (FY 1900).

At the first year, the flow of migrants is highly unequal. The probability to migrate is rather low, except for the very top types, the small minority of individuals who would migrate almost as soon as the option to do so becomes available. However, the accumulated probability to migrate over the entire period of ages 20–30, represented by the left curve in Figure 12, a larger share of the population is likely to participate in the migration movement. Nevertheless, more than half of the total population are virtual “stayers” and are very unlikely to ever migrate.

Figure 13 provides a sense of how the types are distributed across districts. The districts are grouped by deciles of mean type, where higher ranges represent groups of districts with higher mean types. Each range maps the type percentiles within a group of districts onto the Pale-wide type percentiles. The bottom decile of districts (D1) is for the most part excluding itself from participation in the migration movement. Less than 10 percent of individuals in this group are mapped onto the 60th percentile and above of Pale-wide types; below this range the probability to ever migrate is infinitesimal. However, higher deciles have significant shares of potential migrants. Even at D3, more than 20 percent of the individuals are mapped to above the 60th Pale-wide percentile. At the top decile (D10), this rate goes up to almost 70 percent, where the top 20 percent within this group have a probability of at least 80 percent to migrate between ages 20–30, if linked.

Nevertheless, in all districts there are substantial proportions of virtual stayers. Even at the top decile, 30 percent of individuals are mapped below the 60th Pale-wide percentile, meaning that they are not potential migrants. These patterns altogether suggest that both the within-district variation of types, as well as the across district variation in mean types, play important roles in determining the rates of migration.⁷⁸

⁷⁸ The estimates on which these calculations are based do not take into account the various coverage rates in the measurement of migration. Lower coverage rate for a district is likely to bias downwards its estimated type, and thus the across-district variation in mean types is likely to appear greater than it actually is.

6.5.2 Persistence of Short Term Shocks

The distribution of unobserved types is a crucial factor determining the degree to which short term income shocks persist over the long run. The top types are not going to be strongly affected by a business cycle shock in the long run: for individuals at the 99th Pale-wide percentile, the rate of migration at age 20 in a year in which the US business cycle shock is zero (i.e., income is on trend) is 27.4 percents (see Figure 12). In the worst 2.5 percent of the business cycle shocks it will go as low as 22 percent, such that an extreme “once in 40 years shock” would prevent the migration of 19.6 percent of the would be migrants of this type at that year. Those affected by the shock would almost surely migrate at a later year before they reach 30. Thus, a temporary shock may affect them only in the short run; in the long run they are all migrants. The 70-95 percentiles are likely, but are not certain, to migrate. A once in a 40 years shock would reduce the rate of migration at the 90th percentile from 5.6 percent to 4 percent (a reduction of 28.4 percent), and there is no guarantee that these affected would-be migrants would migrate instead any time in the following ten years. The lower the type—the more likely is a temporary shock to persist.

To get a more concrete sense of the persistence of short term income shocks, Plots A.i. and A.ii. of Figure 14 present a simulation exercise. The simulation is over a population that is 20 years old at FY 1900, reflecting the Pale at large in terms of the types distribution. At that year, they are all assumed to be linked and present; that is, they have not previously migrated and they have the option to do so if they wish. The simulation follows income shocks of four different magnitudes, all occurring at FY 1900, of -2, -1, 1, and 2 standard deviations relative to the US income trend ($\sigma = 0.0225$ log points). The migration probabilities of individuals during the 11 years until age 30 (FY 1900–1910) were calculated for a battery of simulated paths of AR(1) shocks that determine the yearly deviations from the income trend. Thus, the first year income shocks persisted through the following years while gradually decaying.⁷⁹

Plot A.i. of Figure 14 reports for each year how much greater or smaller was the number of migrants relative to a scenario of a zero-shock at the first year. For example, a 2 standard deviations shock brought 32.9 percent more immigrants at the same year. The persistence of the income shock made migration in the following two years greater as well, only to subside at the fourth year (age 23/FY 1903). Thereafter, the rates of migration were consistently below those of the zero-shock case. For example, at age 30 (FY 1910), there were still 3.84 percent less migrants compared to the zero-shock scenario.⁸⁰ This long term decline in the rates of migration is attributed to the decimation of the cohorts of individuals that are likely to migrate, caused by the positive shock of the earlier years. Many individuals would not have migrated early if it was not for the positive shock, but some of them would have migrated anyway in one of the later years. This pattern repeats itself, in virtual

⁷⁹ For example, in the case of a 1 standard deviation shock in FY 1900 (equivalent to 2.28 percent greater than the income trend), the expected income in the second year (FY 1901) would be 1.44 percent above trend (based on the estimated root of 0.634).

⁸⁰ Since the rate of migration is declining over the years, a given percentage difference from the no-shock scenario would become smaller in absolute terms.

mirror image, in the case of a negative shock: The decline in migration during the first three years, in which the income shock dominated, was compensated by delayed migration later on, similar to the dynamics illustrated in Figure 3.

Plot A.ii. of Figure 14 quantifies the long-run offsetting of the short-term effect of the business cycle shock. In each scenario at each year, the cumulative number of additional immigrants is added up. In the figure this number is normalized to equal 1 at the year in which this total sum is greatest—in all cases, this happens at age 23 (FY 1903), as seen by the top of the humps—such that the difference in the number of immigrants in the first four years is regarded as the short term effect of the business cycle shocks. From the fifth year on (age 24/FY 1904), the difference in the rates of migration is reversed, such that the short term effect is being offset. By the eleventh year (age 30/FY 1900), more than 40 percent of the short-term effect of the shock has been offset, and there are virtually no differences in this rate across the various simulated shocks, both negative and positive. Put simply, out of six individuals that are affected in the first three years of a business cycle shock, only six remain affected in the long-run of 11 years (and the longer the horizon, even fewer remain affected.)

Since the income shock at the first year has persistence, the rates of migration in the period FY 1901-1911 reflect two competing effects: the persistence of the income shock and the decimation of the cohorts. The first effect reinforces the initial income shock, whereas the second effect offsets it. Due to the persistence effect, the conditional probability to migrate changes throughout the period in the direction of the first year shock. An alternative way to think of the aftermath of an income shock is to separate the two effects and look only at the latter. In plots B.i. and B.ii. of Figure 14. the curves describe a simulation of a shock followed by a “reset”: An income shock takes place in FY 1900, but the income process continues at FY 1901 as if the income was on trend in FY 1900. This enables to view in isolation how a one-off shock that changes the conditional probabilities to migration in a single period only, is offset in the long run by the change to the cohorts of prospective migrants.

As expected, Plot B.i. shows how, following a positive income shock that is reset and not transmitted to the next period, the rates of migration are on average lower, starting from the second period on. In the case of a positive shock of 2 standard deviations, the contemporary increase in the rate of migration is the same as in A.i. (32.9 percent), but in FY 1901 migration is on average 4.68 percent lower than it would have been had FY 1900 income was on trend. This difference gradually converges to zero over time. By FY 1910, migration is on average 1.47 percent lower than in the no-shock scenario. Plot B.ii. shows how the first year shock is cumulatively offset over the coming years. By FY 1910, it is down to only 28.15 percent of its maximum level. That is, almost 72 percent of the prospective migrants that were affected by the shock in the first year would have migrated some time during the following ten years.

Since these simulation exercises above assume that individuals are all linked, but that none have migrated yet, it provides an upper bound for the degree to which the short term effects are offset

in the long run. As explained above, higher types (“movers”) are less affected in the long run by temporary shocks. If the simulation was conducted over a population that had already been exposed to migration, then the share of movers would have been smaller, and thus the temporary shocks would on average have a greater permanent effect. Furthermore, these simulations do not take into account that the formation of networks is endogenous. In reality, a positive temporary shock to the number of migrants should increase the linkage rate in subsequent years, and thus these simulations do not take into account a potentially important channel through which temporary shocks are transmitted into long run shocks. This channel should be effective when the networks are far from being saturated. When approaching saturation the linkage rate can hardly increase, such that temporary shocks to migration no longer affect the future number of migrants through this channel.

7 Conclusion

Under reconstruction, we appreciate your patience

A Data Appendix

A.1 Sample Definition

A.2 The Matching Process

A.3 Who-is-a-Jew Algorithm

References

- Grda, Cormac (2006), *Jewish Ireland in the Age of Joyce: A Socioeconomic History*, Princeton University Press, 322 pp.
- Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson (2012), *A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration*, Working Paper 18011, National Bureau of Economic Research.
- (2013), “Have the poor always been less likely to migrate? Evidence from inheritance practices during the age of mass migration”, *Journal of Development Economics*, Migration and Development 102.
- (2014), “A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration”, *Journal of Political Economy* 122:3.
- Alroey, Gur (2008), *The Quiet Revolution: Jewish Emigration from the Russian Empire 1875-1924 (in Hebrew)*, Jerusalem: The Zalman Shazar Center for Jewish History, 281 pp.
- Aronson, Irwin M. (1990), *Troubled Waters: The Origins of the 1881 Anti-Jewish Pogroms in Russia*, Pittsburgh, PA: University of Pittsburgh Press, 312 pp.
- Association, Jewish Colonization (1906), *Recueil de Materiaux Sur La Situation Economique Des Israelites de Russie VI (1906)*, Paris: Felix Alcan, 446 pp.
- Baines, Dudley (1994), “European Emigration, 1815-1930: Looking at the Emigration Decision Again”, *The Economic History Review* 47:3.
- (1995), *Emigration from Europe 1815-1930*, Cambridge: Cambridge University Press, 100 pp.
- Bandiera, Oriana, Imran Rasul, and Martina Viarengo (2013), “The Making of Modern America: Migratory Flows in the Age of Mass Migration”, *Journal of Development Economics* 102.
- Baron, Salo W. (1976), *The Russian Jew under Tsars and Soviets*, New York, NY: Macmillan.
- Barro, Robert J. and Jos F. Ursa (2010), *Barro-Ursua Macroeconomic Data Set*, URL: <http://rbarro.com/data-sets/> (visited on 10/24/2012).
- Bass, Frank M. (1969), “A New Product Growth for Model Consumer Durables”, *Management Science* 15:5.
- Bohlin, Jan and Anna-Maria Eurenus (2010), “Why They Moved: Emigration from the Swedish Countryside to the United States, 1881-1910”, *Explorations in Economic History* 47:4.
- Borjas, George J. (1987), “Self-Selection and the Earnings of Immigrants”, *The American Economic Review* 77:4.
- Boswijk, H. Peter and Philip Hans Franses (2005), “On the Econometrics of the Bass Diffusion Model”, *Journal of Business & Economic Statistics* 23:3.
- Boustan, Leah Platt (2007), “Were Jews Political Refugees or Economic Migrants? Assessing the Persecution Theory of Jewish Emigration, 1881?1914”, in *The New Comparative Economic History: Essays in Honor of Jeffrey G. Williamson*, ed. by Timothy J. Hatton, Kevin H. O’Rourke, and Alan M. Taylor, Cambridge, Mass: MIT Press.
- Brutskus, Boris D. (1909), *Statistika evreiskago naseleniia : raspredelenie po territorii, demograficheskie i kul’turnye priznaki evreiskago naseleniia po dannym perepisi 1897 g.* St. Peterburg: Evreiskoe Kolonizatsionnoe obshchestvo.
- Chiswick, Barry R. (1991), “Jewish Immigrant Skill and Occupational Attainment at the Turn of the Century”, *Explorations in Economic History* 28:1.
- (1992), “Jewish Immigrant Wages in America in 1909: An Analysis of the illingham commission data”, *Explorations in Economic History* 29:3.

- Clark, Ximena, Timothy J Hatton, and Jeffrey G Williamson (2007), “Explaining U.S. Immigration, 1971?1998”, *Review of Economics and Statistics* 89:2.
- Clem, Ralph S., ed. (1986), *Research Guide to the Russian and Soviet Censuses*, Studies in Soviet History and Society, Ithaca: Cornell University Press, 323 pp.
- Committee, Russo-Jewish (1890), *The persecution of the Jews in Russia*, London: Wertheimer, Lea & Co., 102 pp.
- Davis, Joseph H. (2006), “An Improved Annual Chronology of U.S. Business Cycles since the 1790s”, *The Journal of Economic History* 66:1.
- Deltas, George, Richard Sicotte, and Peter Tomczak (2008), “Passenger Shipping Cartels and Their Effect on Trans-Atlantic Migration”, *Review of Economics and Statistics* 90:1.
- Diner, Hasia (1995), “Before the Promised City: Eastern European Jews in America Before 1880”, in *An Inventory of Promises: Essays on American Jewish History in Honor of Moses Rischin*, ed. by Jeffrey Gurock and Marc L. Raphael, New York, NY: Carlson.
- Dubnow, Simon (1916), *History of the Jews in Russia and Poland, from the Earliest Times Until the Present Day*, trans. by Israel Friedlaender, Philadelphia: Jewish Publication Society of America.
- Easterlin, Richard A. (1961), “Influences in European Overseas Emigration before World War I”, *Economic Development and Cultural Change* 9:3.
- Eckstein, Zvi and Yoram Weiss (2004), “On the Wage Growth of Immigrants: Israel, 1990?2000”, *Journal of the European Economic Association* 2:4.
- Ferenczi, Imre and Walter F. Willcox (1929), *International Migrations*, vol. I: Statistics, 2 vols., New York, NY: National Bureau of Economic Research.
- Frankel, Jonathan (1981), *Prophecy and Politics: Socialism, Nationalism, and the Russian Jews, 1862-1917*, New York, NY: Cambridge University Press, 716 pp.
- Gallant, A. Ronald, Han Hong, and Ahmed Khawaja (2010), “Estimating a Dynamic Oligopolistic Game with Serially Correlated Unobserved Production Costs”, unpublished manuscript.
- Godley, Andrew (2001), *Jewish Immigrant Entrepreneurship in New York and London 1880-1914*, New York, NY: Palgrave Macmillan.
- Goldsmith, Raymond W. (1961), “The Economic Growth of Tsarist Russia 1860-1913”, *Economic Development and Cultural Change* 9:3.
- Gould, Eric D. and Omer Moav (2014), “Does High Inequality Attract High Skilled Immigrants?”, *The Economic Journal*.
- Gould, John D. (1979), “European Inter-Continental Emigration 1815-1914: Patterns and Causes.”, *Journal of European Economic History* 8:3.
- (1980), “European Inter-Continental Emigration: The Role of 'Diffusion' and 'Feedback'”, *Journal of European Economic History* 9:2.
- Gregory, Paul R. (1982), *Russian National Income, 1885-1913*, Cambridge and New-York, NY: Cambridge University Press, 359 pp.
- Hatton, Timothy J. (1995), “A Model of U.K. Emigration, 1870-1913”, *The Review of Economics and Statistics* 77:3.
- (2010), “The Cliometrics of International Migration: A Survey”, *Journal of Economic Surveys* 24:5.
- Hatton, Timothy J. and Jeffrey G. Williamson (1998), *The Age of Mass Migration: Causes and Economic Impact*, New York, NY: Oxford University Press, 318 pp.
- (2008), *Global Migration and the World Economy: Two Centuries of Policy and Performance*, Cambridge, Mass: MIT Press, 488 pp.

- Hersch, Liebmann (1913), *Le Juif Errant d'Aujourd'hui*, Paris: Giard & Briere, 368 pp.
- Jerome, Harry (1926), *Migration and Business Cycles*, New York, NY: National Bureau of Economic Research, 266 pp.
- JewishGen, *JewishGen Communities Database*, URL: <http://www.jewishgen.org/communities/search.asp> (visited on 10/31/2012).
- Johnpoll, Bernard K. (1995), "Why They Left: Russian-Jewish Mass Migration and Repressive Laws, 1881-1917", *American Jewish Archives* 47:1.
- Joseph, Samuel (1914), *Jewish Immigration to the United States from 1881 to 1910*, New York, NY: Columbia University, 216 pp.
- Kahan, Arcadius (1986a), *Essays in Jewish Social and Economic History*, Chicago, IL & London: University of Chicago Press.
- (1986b), "The Urbanization Process of the Jews in Nineteenth-Century Europe", in *Essays in Jewish Social and Economic History*, ed. by Roger Weiss, Chicago, IL & London: University of Chicago Press.
- Karemera, David, Victor Iwuagwu Oguledo, and Bobby Davis (2000), "A gravity model analysis of international migration to North America", *Applied Economics* 32:13.
- Keane, Michael P. and Robert M. Sauer (2009), "Classification Error in Dynamic Discrete Choice Models: Implications for Female Labor Supply Behavior", *Econometrica* 77:3.
- Keeling, Drew (2011), "North Atlantic Shipping Cartels and the effects of the 1904 Fare War upon Migration between Europe and the United States", in *Regulierte Mrkte: Znfte und Kartelle*, ed. by Magrit Mller, Heinrich R. Schmidt, and Tissot Laurent, Chronos, vol. 26, Schweizerische Gesellschaft fr Wirtschafts- und Sozialgeschichte, Zrich.
- Klier, John D. (1986), *Russia Gathers Her Jews: The Origins of the "Jewish Question" in Russia, 1772-1825*, Dekalb, IL: Northern Illinois University Press, 236 pp.
- (1992), "The Pogrom Paradigm in Russian History", in *Pogroms: Anti-Jewish Violence in Modern Russian History*, ed. by John D. Klier and Shlomo Lambroza, Cambridge and New-York, NY: Cambridge University Press.
- (1995), *Imperial Russia's Jewish Question, 1855-1881*, Cambridge and New-York, NY: Cambridge University Press, 562 pp.
- (2000), "What Exactly was a Shtetl?", in *The Shtetl: Image and Reality*, ed. by Gennady Estraiikh and Mikhail Krutikov, Papers on the Second Mendel Friedman International Conference on Yiddish (Studies in Yiddish), Oxford: Legenda.
- (2011), *Russians, Jews, and the Pogroms of 1881-1882*, Cambridge: Cambridge University Press, 514 pp.
- Kuznets, Simon (1975), "Immigration of Russian Jews to the United States: Background and Structure", *Perspectives in American History* 9.
- Lambroza, Shlomo (1981), "The Pogrom Movement in Tsarist Russia 1903-1906", Unpublished PhD Dissertation, Rutgers University, unpublished dissertation, Rutgers University, New Brunswick, NJ.
- (1992), "The Pogroms of 1903-1906", in *Pogroms: Anti-Jewish Violence in Modern Russian History*, Cambridge: Cambridge University Press.
- Lederhendler, Eli (2009), *Jewish Immigrants and American Capitalism, 1880-1920: From Caste to Class*, 1st ed., Cambridge: Cambridge University Press.
- Lestschinsky, Jacob (1961), "Jewish Migrations, 1840-1956", in *The Jews: their History, Culture and Religion*, ed. by Louis Finkelstein, 3rd ed., vol. 2, New-York, NY: Harper & Row Publishers New-York.

- Mayda, Anna Maria (2009), “International migration: a panel data analysis of the determinants of bilateral flows”, *Journal of Population Economics*.
- McKenzie, David and Hillel Rapoport (2010), “Self-selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks”, *Review of Economics and Statistics* 92:4.
- Mendelsohn, Ezra (1970), *Class Struggle in the Pale: The Formative Years of the Jewish Worker’s Movement in Tsarist Russia*, 1st ed., Cambridge: Cambridge University Press, 196 pp.
- Milamed, Susan (1986), “Proskurover Landsmanshaftn: A Case Study in Jewish Communal Development”, *American Jewish History* 76:1.
- Moretti, Enrico (1999), “Social Networks and Migrations: Italy 1876-1913”, *International Migration Review* 33:3.
- Motzkin, Leo (1910), *Die Judenpogrome in Russland*, Berlin: Jdischer Verlag, 436 pp.
- Moya, Jose C. (2005), “Immigrants and Associations: A Global and Historical Perspective”, *Journal of Ethnic & Migration Studies* 31:5.
- Munshi, Kaivan (2003), “Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market”, *The Quarterly Journal of Economics* 118:2.
- Nathans, Benjamin (2002), *Beyond the pale: the Jewish encounter with late imperial Russia*, Berkeley & Los Angeles: University of California Press, 448 pp.
- Nelson, Charles R. and Charles R. Plosser (1982), “Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications”, *Journal of Monetary Economics* 10:2.
- O’Rourke, Kevin H. and Jeffrey G. Williamson (1997), “Around the European periphery 1870?1913: Globalization, schooling and growth”, *European Review of Economic History* 1:2.
- Pedersen, Peder J., Mariola Pytlikova, and Nina Smith (2008), “Selection and Network Effects: Migration Flows into OECD Countries 1990?2000”, *European Economic Review* 52:7.
- Perlmann, Joel (2000), “What the Jews Brought: East European Jewish Immigration to the United States, c. 1900”, in *Immigrants, Schooling, and Social Mobility: Does Culture Make a Difference?*, ed. by Hans Vermeulen and Joel Perlmann, New York, NY & London: Palgrave Macmillan.
- (2001), “Race or People: Federal Race Classifications for Europeans in America, 1898–1913”, *Jerome Levy Economics Institute Working Paper No. 320*.
- (2006), “The Local Geographic Origins of Russian-Jewish Immigrants, Circa 1900”, *Levy Economics Institute Working Paper No. 465*.
- Petrovsky-Shtern, Yohanan (2008), *Jews in the Russian army, 1827-1917: Drafted into Modernity*, Cambridge and New-York, NY: Cambridge University Press, 1 p.
- Rogger, Hans (1986), *Jewish Policies and Right-Wing politics in Imperial Russia*, Berkeley and Los-Angeles, CA: University of California Press.
- Rowland, Richard H. (1986a), “Geographical Patterns of the Jewish Population in the Pale of Settlement of Late Nineteenth Century Russia”, *Jewish Social Studies* 48:3.
- (1986b), “Urbanization and Migration Data in Russian and Soviet Censuses”, in *Research Guide to the Russian and Soviet Censuses*, Ithaca, NY: Cornell University Press.
- Rubinow, Isaac M. (1907), “Economic Condition of the Jews in Russia”, *U.S. Bureau of Labor Bulletin* 15.
- Ruppin, Arthur (1934), *The Jews in the Modern World*, London: Macmillan and Co., 466 pp.
- Rust, John (1987), “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher”, *Econometrica* 55:5.
- Sarna, Jonathan D. (1981), “The myth of no return: Jewish return migration to Eastern Europe, 1881-1914”, *American Jewish History* 71:2.

- “Jewish National Organizations in the United States” (1919), *American Jewish Year Book* 21, ed. by Harry Schneiderman.
- Slezkine, Yuri (2004), *The Jewish Century*, Princeton, NJ: Princeton University Press, 462 pp.
- Snchez-Alonso, Blanca (2000), “Those Who Left and Those Who Stayed Behind: Explaining Emigration from the Regions of Spain, 1880?1914”, *The Journal of Economic History* 60:3.
- Soyer, Daniel (1997), *Jewish Immigrant Associations and American Identity in New York, 1880-1939*, Cambridge, Mass: Harvard University Press.
- Spitzer, Yannay (2014), “Pogroms Networks and Migration: The Jewish Migration from the Russian Empire to the United States 1881–1914”, unpublished manuscript.
- Spitzer, Yannay and Ariell Zimran (2014), “Migrant Self-Selection: Anthropometric Evidence from the Mass Migration of Italians to the United States, 1907?1925”, unpublished manuscript.
- Srinivasan, V. and Charlotte H. Mason (1986), “Technical Note: Nonlinear Least Squares Estimation of New Product Diffusion Models”, *Marketing Science* 5:2.
- Stampfer, Shaul (1986), “The Geographic Background of East European Jewish Migration to the United States before World War I”, in *Migration Across Time and Nations: Population Mobility in Historical Contexts*, ed. by Ira A. Glazier, New York: Holmes and Meier.
- (1989), “The 1764 Census of Polish Jewry”, *Annual of Bar-Ilan University Studies in Judaica and the Humanities* 24.
- Szajkowski, Zosa (1942), “How the Mass Migration to America Began”, *Jewish Social Studies* 4:4.
- Szold, Henrietta (1906), “From Kishineff to Bialystok: A Table of Pogroms from 1903 to 1906”, *American Jewish Year Book* 8.
- Thomas, Brinely (1954), *Migration and Economic Growth*, Cambridge: Cambridge University Press.
- Thomas, Dorothy S. (1941), *Social and Economic Aspects of Swedish Population Movements, 1750-1933*, New York, NY: The Macmillan Company, 520 pp.
- Tsentralnyi Statisticheskii Komitet, [Central Statistical Committee] (1905), *Naselennia mesta Rossiiskoi Imperii v 500 i bolee zhitelei s ukazaniem vsego nalichnago v nikh naseleniia i chisla zhitelei preobladaiushchikh veroispovedanii, po dannym Pervoi vseobshchei perepisi naseleniia 1897 g.* St. Peterburg: Tip. Obshchestvennaia polza.
- Weber, John B. and Walter Kempster (1892), *Report of the Commissioners of Immigration Upon the Causes which Incite Immigration to the United States*, Washington, DC: Government Printing Office, 334 pp.
- Wegge, Simone A. (1998), “Chain Migration and Information Networks: Evidence from Nineteenth-Century Hesse-Cassel”, *The Journal of Economic History* 58:4.
- Weil, Patrick (2000), “Races at the Gate: A Century of Racial Distinctions in American Immigration Policy (1865-1965)”, *Georgetown Immigration Law Journal* 15.

Table 1: Variables and Parameters in the Migration Model

A. VARIABLES

<i>Applies at</i>	<i>Notation</i>	<i>Description</i>
Macro	\tilde{w}_t	Income trends
	d_t	Deviations from income trends
Individual	η_i	Individual type
	ζ_{it}	Individual cycle
	ξ_{it}	Individual transitory term
	$l_{i,t-1}$	Location at the beginning of period t
	l_{it}	Location of choice
	a_{it}	Age
	k_{it}	Linkage status
District	η_j	Mean district type
	x_j	District characteristics
	ϵ_j	District random-effect
	n_{jt}	Recent network activity
	N_{jt}	Size of effective network
	K_{jt}	Linkage rate

B. PARAMETERS

<i>Applies at</i>	<i>Notation</i>	<i>Description</i>
Macro	g	Growth rate of income trends
	φ	Roots of B.C. AR(1) processes
	σ_μ	Variances of AR(1) random terms
Individual	β	Discount rate
	γ	Parameters of cost function
	κ	Coeffs. (time, exper., yields)
	ζ	Amplitude, individual cycle
	p	Transition prob., individual cycle
	σ_ξ	Variance of transitory term
District	σ_η	Variance of types within districts
	σ_ϵ	Variance of types across districts
	α	Mean types function
	δ, λ	Linkage function

Table 2: Descriptive Statistics

<i>A. District</i> (N=215)	Average	St.Dev.	Min.	Q1	Median	Q3	Max.
Migration/k (raw)	44.69	44.28	2.03	18.24	33.80	57.46	306.47
Migration/k	147.10	151.28	6.00	57.68	110.65	190.06	1078.12
Jewish population (1,000s)	21.75	21.21	1.58	9.18	16.09	27.70	198.99
Total population (1,000s)	181.45	109.15	43.70	100.66	156.71	233.56	826.78
Number of towns identified	1.98	1.23	1.00	1.00	2.00	3.00	6.00
Coverage	0.59	0.26	0.13	0.44	0.58	0.71	1.97
Commerce-to-manufacturing	-0.13	0.93	-2.58	-0.77	-0.02	0.61	1.93
<i>B. District × Year</i> (N=3,225)							
Migration/k (raw)	4.19	5.91	0.00	1.07	2.67	5.41	151.46
Migration/k	13.98	20.57	0.00	3.86	8.92	17.62	608.89
Migration/k, age 16-20	20.75	29.20	0.00	4.61	12.17	26.93	580.27
Migration/k, age 46-50	6.92	13.19	0.00	0.00	3.08	9.60	402.01
Effective network (growth)	1.05	3.21	0.00	0.00	0.00	0.00	63.45
Effective network (stock)	4.75	6.92	0.00	0.00	2.06	6.85	63.45
<i>C. Cohort × District</i> (N=10,535)							
Migration/k (raw)	39.75	53.90	0.00	6.60	24.63	52.63	584.80
Migration/k	132.51	184.38	0.00	22.67	81.44	174.26	2194.68
Migration/k, 1863 Cohort	96.25	132.65	0.00	21.16	60.75	123.20	969.30
Migration/k, 1884 Cohort	236.78	259.74	0.00	91.25	161.76	300.87	1759.32
<i>D. Cohort × District × Year</i> (N=158,025)							
Migration/k (raw)	3.71	7.67	0.00	0.00	0.00	4.91	275.36
Migration/k	12.37	26.47	0.00	0.00	0.00	16.34	1106.99

Note: The sample includes all migrations records whose last place of residence was identified, and were within the sample age interval (16-50) in the respective year. All migration counts are per 1,000 Jews in the respective cohorts as of the 1897 census. Unless specified as "raw", the measure is adjusted by inflating the number of immigrants in each year according to the share of total to observed migration. Coverage is the share of the Jewish population in the identified towns to the total Jewish population in the district; it may exceed 1 since the Jewish population in the towns was defined by religion, whereas the districts population by mother language. Commerce-to-manufacturing is the log of the share of Jews employed in commerce to Jews employed in manufacturing, normalized to have a mean of zero and standard deviation of 1 across the Pale's district (*not* across the current sample). Effective networks are per 100,000 Jews in the district (1897). The stock of effective network is the number of all hometown based associations related to the district incorporated before or during the respective year, weighted by years since incorporation. The growth of the effective network is the number of associations incorporated during that year. The 1863 cohort (aged 50 in 1913) is the oldest cohort that is covered throughout the entire sample period. The 1884 cohort (aged 16 in 1899) is the youngest cohort fully covered.

Table 3: Determinants of migration—reduced form

DEP. VAR.:	mig./k (mean = 11.76)				log mig./k (mean = 1.93)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Commerce/manufacturing	-2.579 ^a (0.735)	-2.579 ^a (0.735)	-0.431 (0.894)		-0.210 ^a (0.074)	-0.210 ^a (0.074)	-0.067 (0.106)	
Associations 1861–1899	4.286 ^a (1.190)	4.286 ^a (1.190)	1.549 (1.196)		0.432 ^a (0.113)	0.432 ^a (0.113)	0.207 ^c (0.120)	
Capital district	4.479 ^c (2.355)	4.479 ^c (2.355)	3.942 ^b (1.917)		0.371 ^c (0.208)	0.371 ^c (0.208)	0.263 (0.163)	
Coverage	6.766 ^c (3.570)	6.766 ^c (3.571)	17.475 ^a (3.137)		0.732 (0.485)	0.732 (0.485)	1.981 ^a (0.405)	
US income	61.447 ^a (5.575)	61.049 ^a (5.571)			9.777 ^a (0.644)	9.722 ^a (0.642)		
Russia income		16.446 ^a (1.661)				2.284 ^a (0.154)		
Yield/seed residual				-1.257 (0.810)				-0.005 (0.092)
Constant	-307.810 (28.394)	-382.371 (30.255)			-48.593 (3.291)	-58.948 (3.663)		
Year F.E.			Yes	Yes			Yes	Yes
Geographic F.E.			Prov.	Dist.			Prov.	Dist.
R-squared	0.195	0.203	0.463	0.695	0.198	0.212	0.507	0.761
p-value of F-stat.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	3,090	3,090	3,090	3,090	3,090	3,090	3,090	3,090

Significance levels: ^a : $p < 0.01$; ^b : $p < 0.05$; ^c : $p < 0.1$.

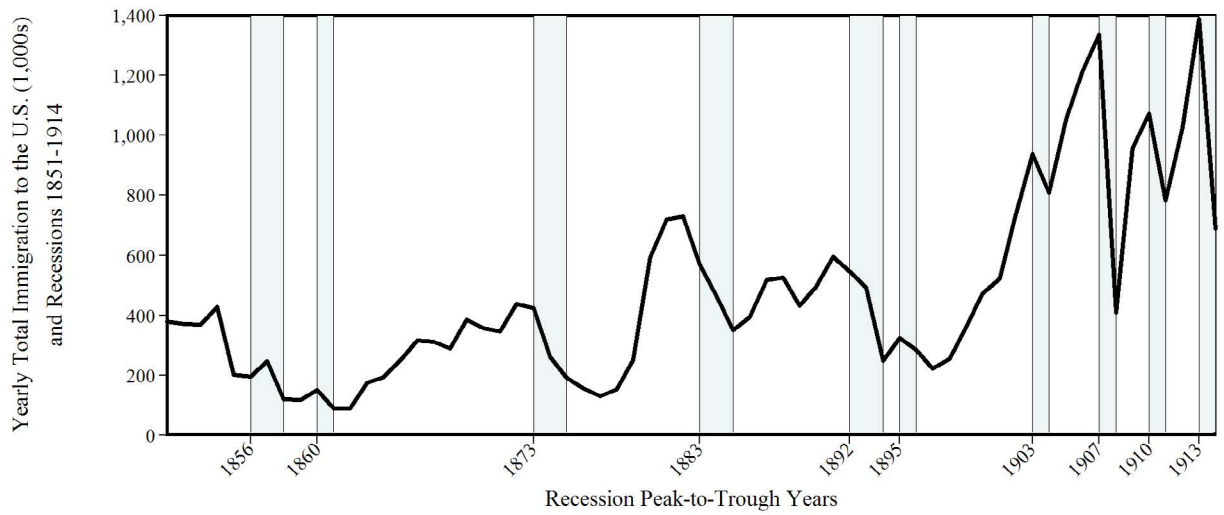
Notes: The table reports OLS regressions predicting migration per 1,000 (ages 16–50), adjusted according to the yearly ratio of observed-to-unobserved migration. Each observation is a year×district, covering the period FY 1900–1914. *Commerce/manufacturing*: standardized log of the ratio of Jews employed in commerce to Jews employed in manufacturing. *Associations*: cumulative number of associations incorporated per 100,000 residents in the district over the period 1861–1899. *Capital district*: an indicator for the principal district of the province. *Coverage*: ratio of Jewish population in towns covered by the geo-matching algorithm to total district population. *US income*: log real wage (Barro and Ursua 2010). *Russia income*: log NNPPC (Gregory 1982). *Yield/seed*: province-year residuals of the yield/seed ratio of five crops from a regression on year and province fixed effects, weighted by seeded area. Standard errors, clustered by district, are reported in parentheses.

Table 4: Parameters estimates—benchmark structural model

Parameter	Symbol	(1) Est.	(2) S.E.
<i>Individual DP problem</i>			
Within-dist. type var.	σ_η	0.412	0.010
B.C., US	γ_1	9.897	0.221
Yield/seed residual	κ_y	0.003	0.005
Time trend	κ_t	0.015	0.000
Experience	κ_e	-0.007	0.000
Constant	α	0.000	
Discount factor	β	0.950	
Cost of migration	γ_0	0.000	
Indiv. cycle amplitude	ζ	0.000	
Indiv. cycle trans. prob.	p	0.000	
District FE		Yes	
<i>Diffusion process</i>			
Depreciation	δ	0.096	0.029
Constant	λ_0	-4.369	0.048
District	λ_d	0.076	0.023
Province	λ_p	3.979	0.077
Log likelihood		-482,171	

Notes: The table reports the estimates of the benchmark model. Standard errors are computed by the inverse of the Hessian. Calibrated parameters are reported without standard errors. The standard deviation of the EVI error term is normalized to $\sigma_\xi = 1$.

(a) Levels of Total U.S. Immigration and Recessions



(b) Changes in Total U.S. Immigration and Recessions

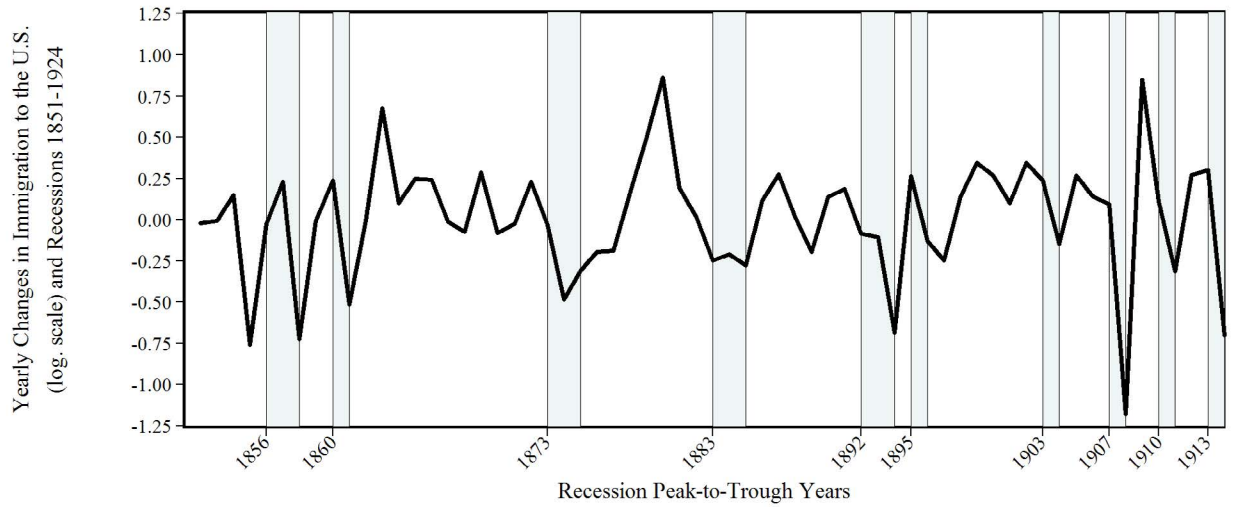


Figure 1: Total U.S. Immigration and Recessions

Note: Shaded areas represent peak-to-trough years in recessions reported by Davis (2006). Yearly flows of U.S. immigration from Ferenczi and Willcox (1929, Table IV).

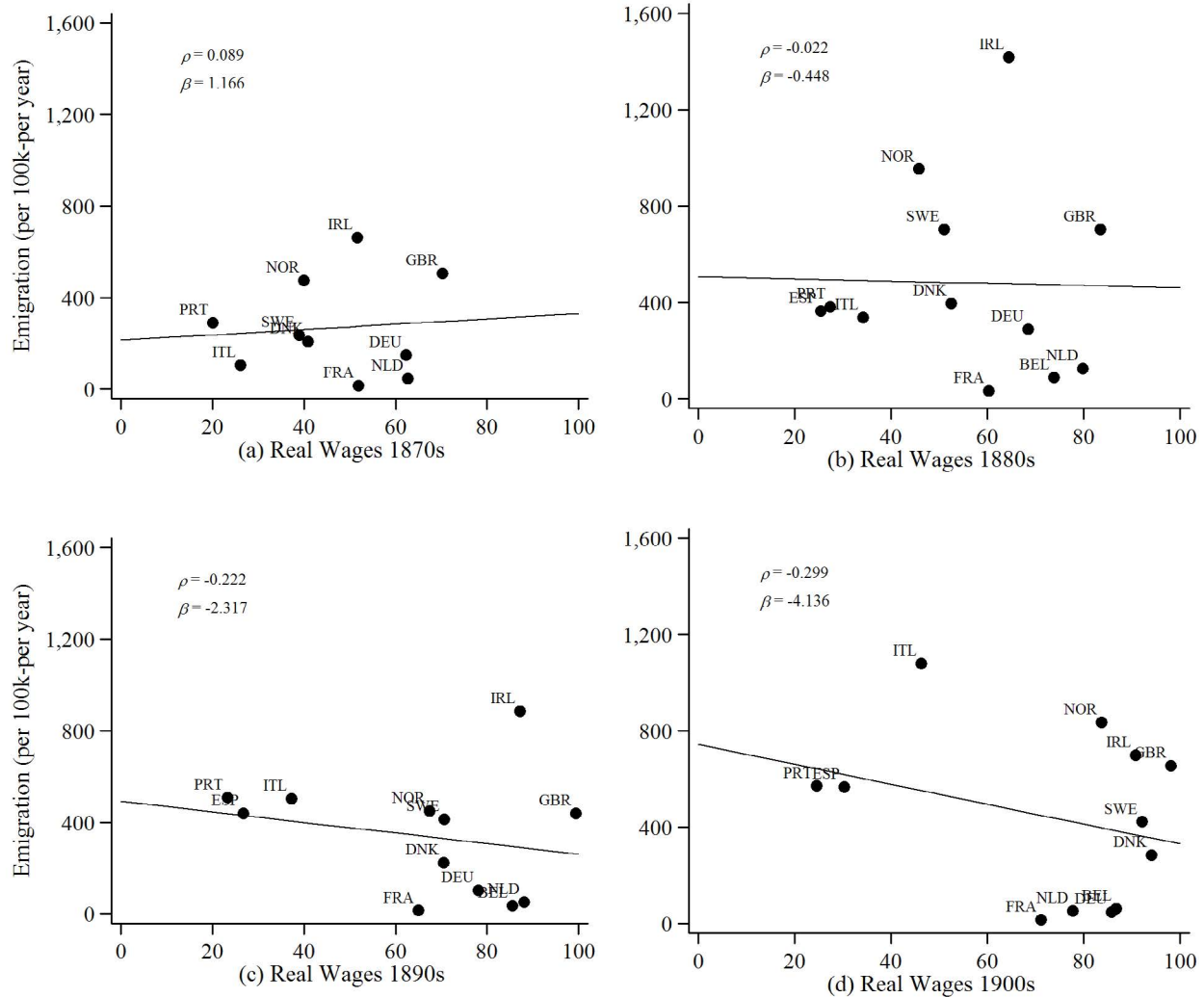


Figure 2: European Emigration and Real Wages 1870-1910

Note: Emigration rates are decade averages in yearly terms per 100,000 (Ferenczi and Willcox 1929, Text Table 9, pp. 200-201). Real wages are internationally comparable PPP-adjusted decade averages revised in O'Rourke and Williamson (1997) as reported in Hatton and Williamson (2008, Table 4.2), where 100 is the level of British real wage in 1905. The real wages are one year lagged relative to migration (e.g., 1870-1879 wages correspond to 1871-1880 emigration). Real wages in the 1900s are for the years 1900-1913. The dotted lines are univariate OLS regression lines.

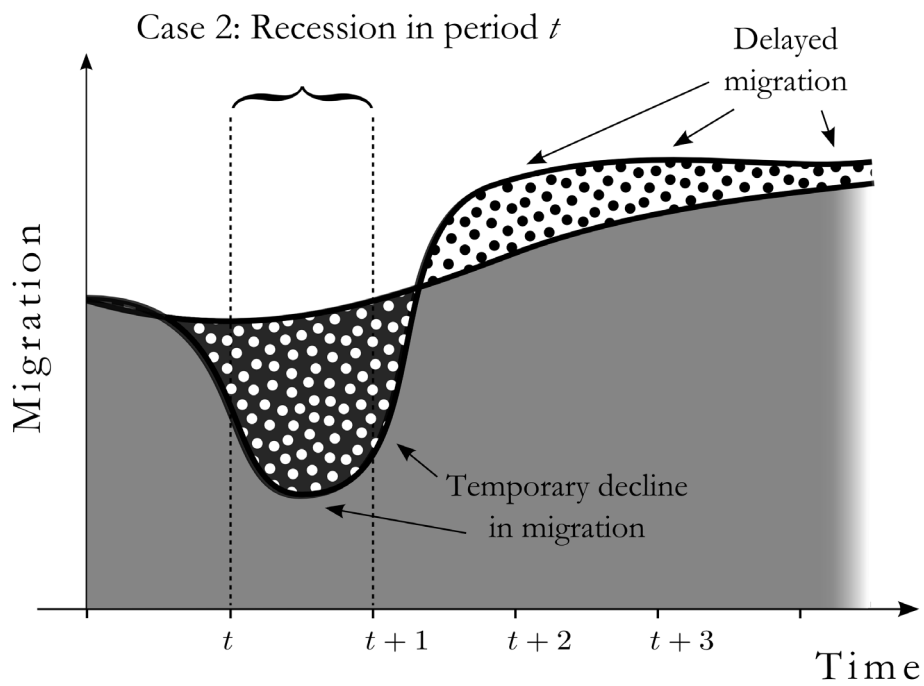
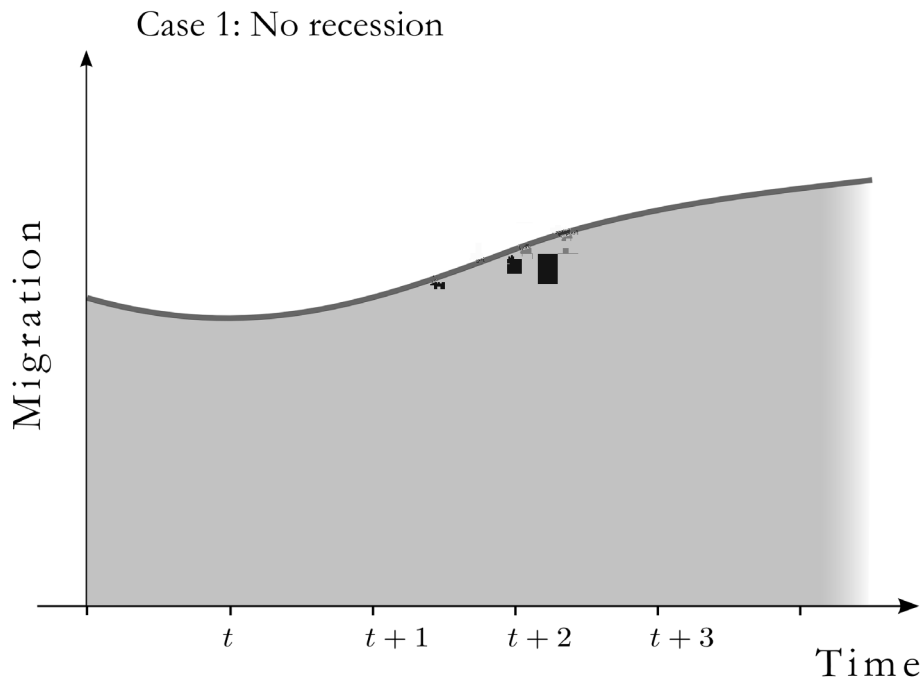
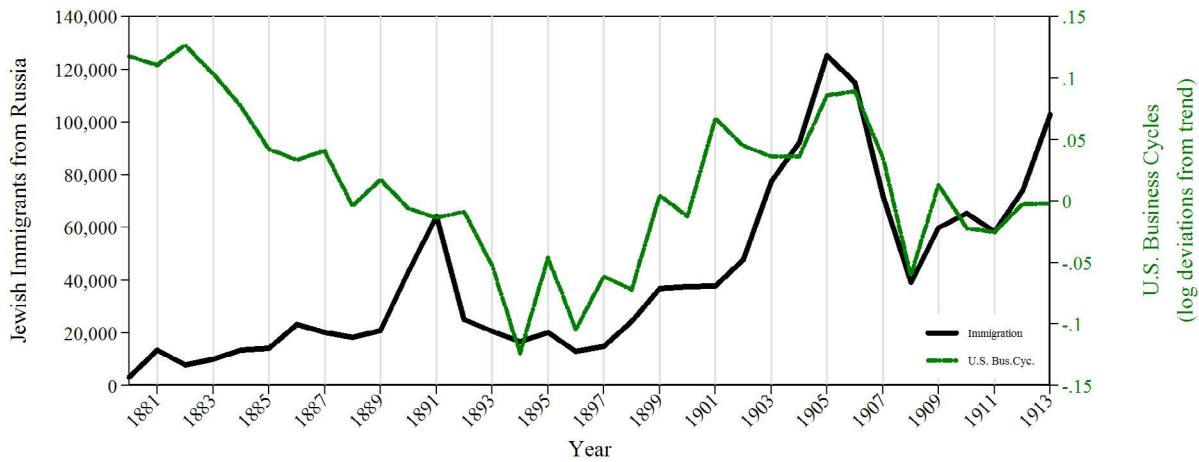


Figure 3: Effect of Temporary ⁶⁹Income Shock on Migration

(a) Jewish Russian Immigration and U.S. Business Cycles



(b) Jewish Russian Immigration and Russian Business Cycles

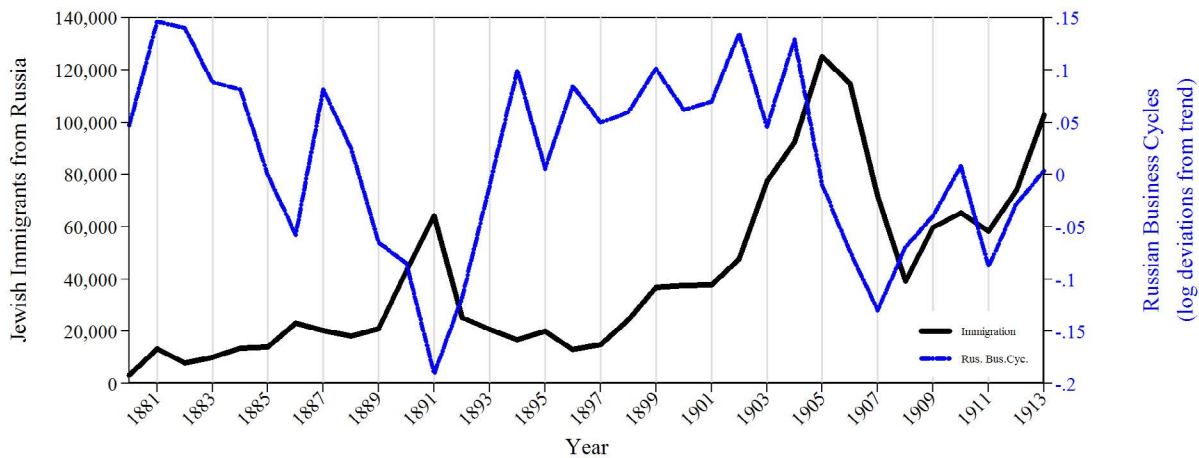


Figure 4: Jewish Immigration from Russia and U.S. Business Cycles

Notes: Jewish Russian immigration from Joseph (1914, Table XII) and Ferenczi and Willcox (1929, Table XXXII), with corrections by Godley (2001, Table 5.4). Russian net national income per-capita from Gregory (1982). U.S. GDP per-capita from Barro and Ursua (2010). Migration years are from July 1st (current year) to June 30th (next year). Migration years started at July 1, e.g., 1880 stands for July 1st, 1880 - June 30th 1881, etc. The data series in the figure goes from July 1st, 1880 to June 30th 1914. The deviations from the trend are the residuals from an OLS regression with AR(1) errors over the period 1885-1913; Russian pre-1885 income data originates from a different series (Barro and Ursua 2010; Goldsmith 1961) and is presented here for completeness.

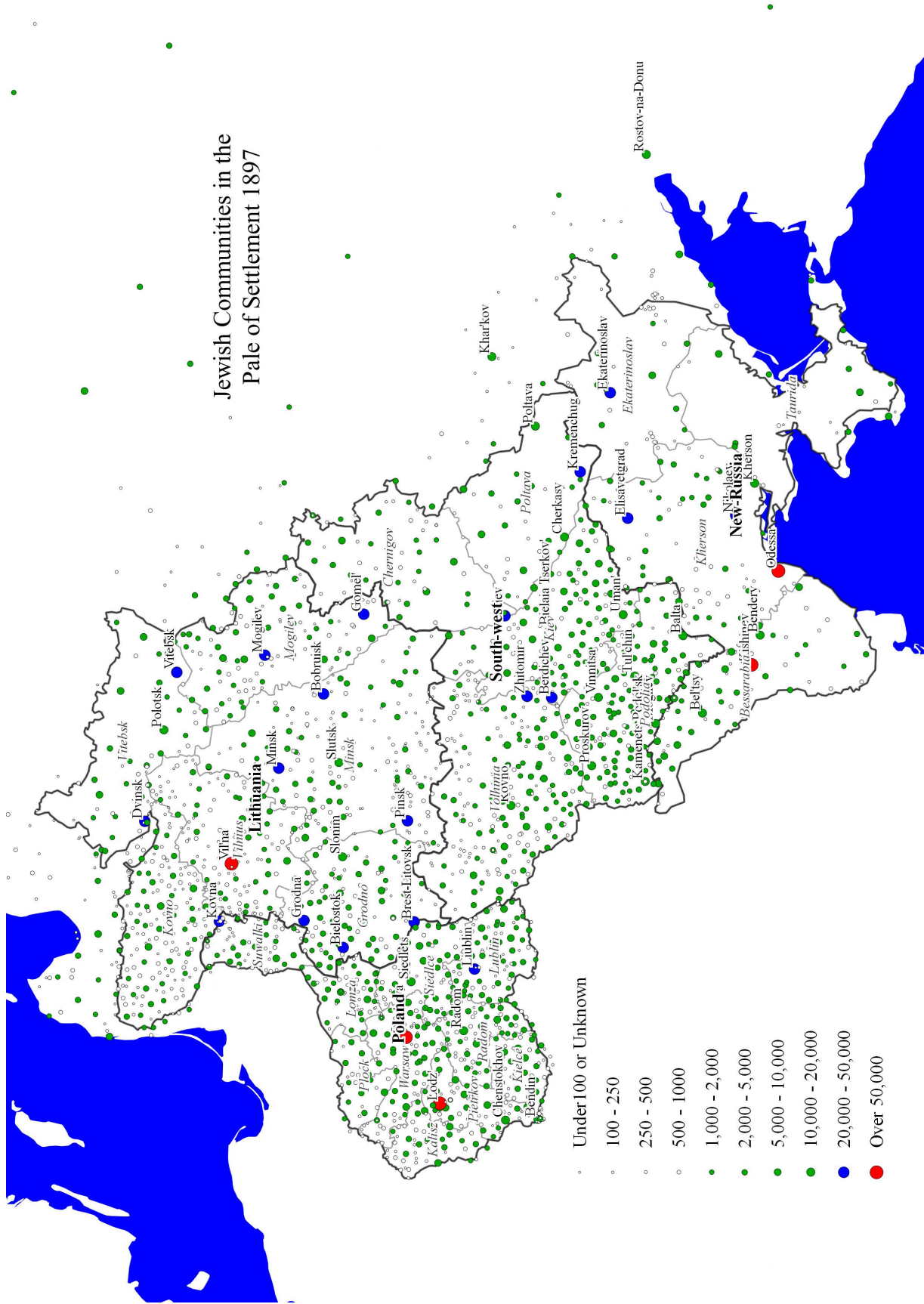


Figure 5: Jewish Communities in the Pale of Settlement 1897

Note: The map includes all localities in the Pale of Settlement in which (a) the total population was at least 500; and (b) The Jewish population was at least 10 percent of the total. The source for the names and the populations is tsentralnyistatisticheskii komitet 1905. Geographic coordinates and additional communities of unknown size, typically of very small localities, have been added from JewishGen ([JewishGen Communities Database](#)). Remaining coordinates were added using various online sources.

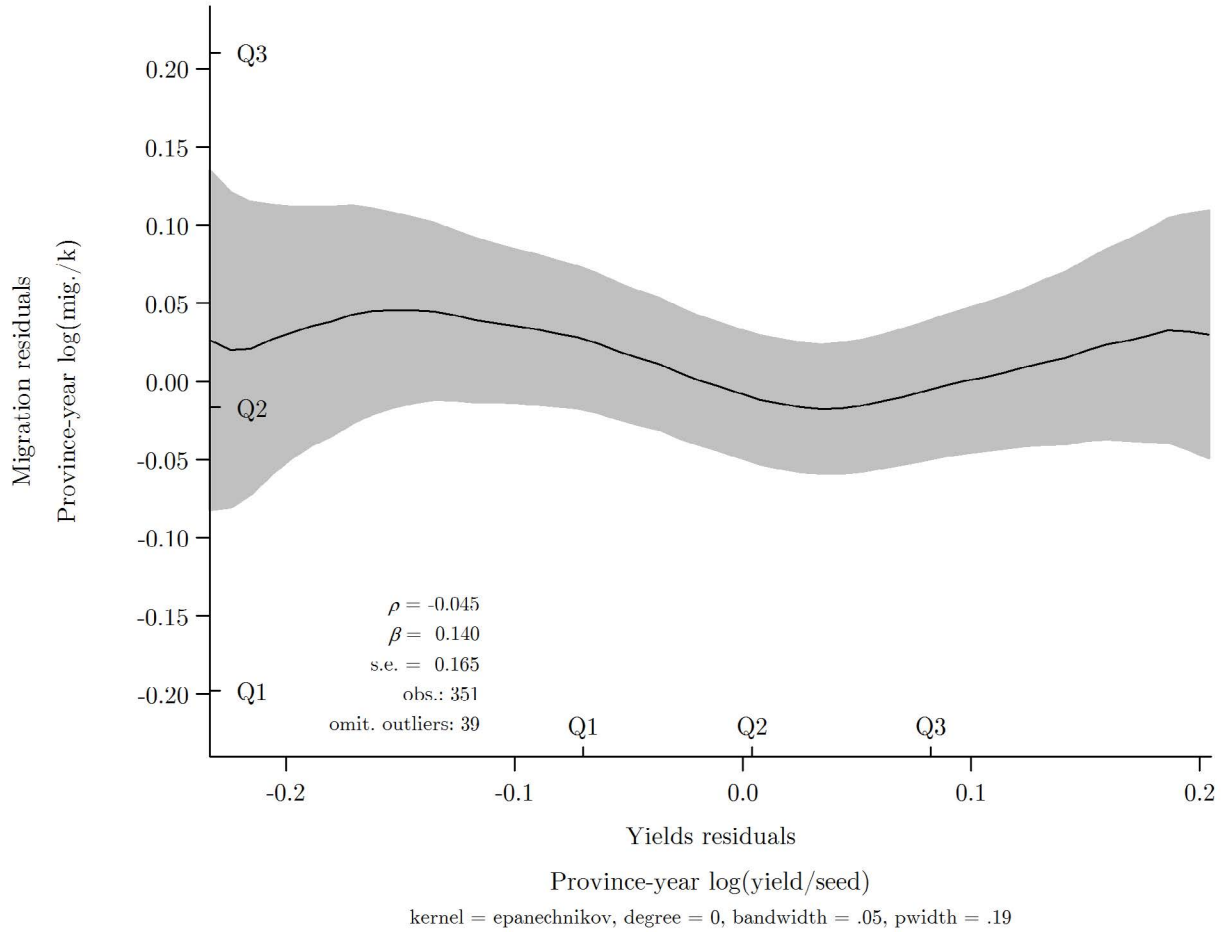


Figure 6: Migration and yield shocks—nonparametric residuals regression

Notes: The plot represents a nonparametric residuals regression of migration on yields shock. Each observation is a province-year, covering 26 provinces and the periods FY 1900–1914 (migration) and 1899–1913 (yields). Migration is measured as the log of (adjusted) migration per thousand, taking into account at each year the cohorts aged 16–50. The migration residuals are taken from a regression of migration on province fixed-effects and year fixed-effects. Yields output are yield/seed ratio of winter rye, spring wheat, barley, oats, and potatoes. The yield residuals are taken from a regression of yield/seed ratio on province fixed-effects and year fixed-effects, separately for each crop, and then averaged using the seeded areas of each crop as weights. The top and bottom 5 percentiles of yield residuals are omitted from the regression. Inner axis marks indicate the quartiles of each of the two variables. The regression uses an Epanechnikov kernel function with bandwidth 0.05. The shaded area represents the 95 percent confidence interval.

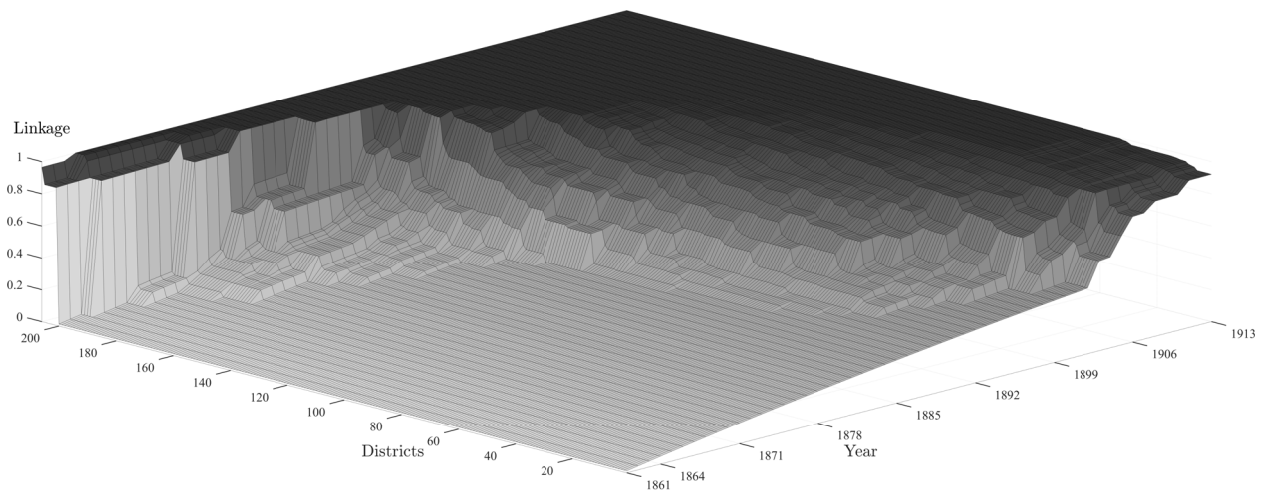


Figure 7: Diffusion of migration networks—estimated CDF of linkage over districts by years

Notes: The figure represents the cumulative distribution function of the estimated linkage rates across districts for each year. Note that districts are ordered by linkage at each year, and their order may change from one year to another. Linkage is the share of individuals born in the district that have already received an option to migrate, including those that have already migrated (e.g., $\text{linkage}_{jt} = 0.6$ implies that the probability of an individual born in district j to not have been linked by year t is 0.4). Linkage is calculated at each year according to the histories of landsmanshaftn incorporation within the district and the province (see section 4.5 for full specification). The parameters of the diffusion process are estimated within the benchmark estimation model, where Russian business cycles are suppressed.

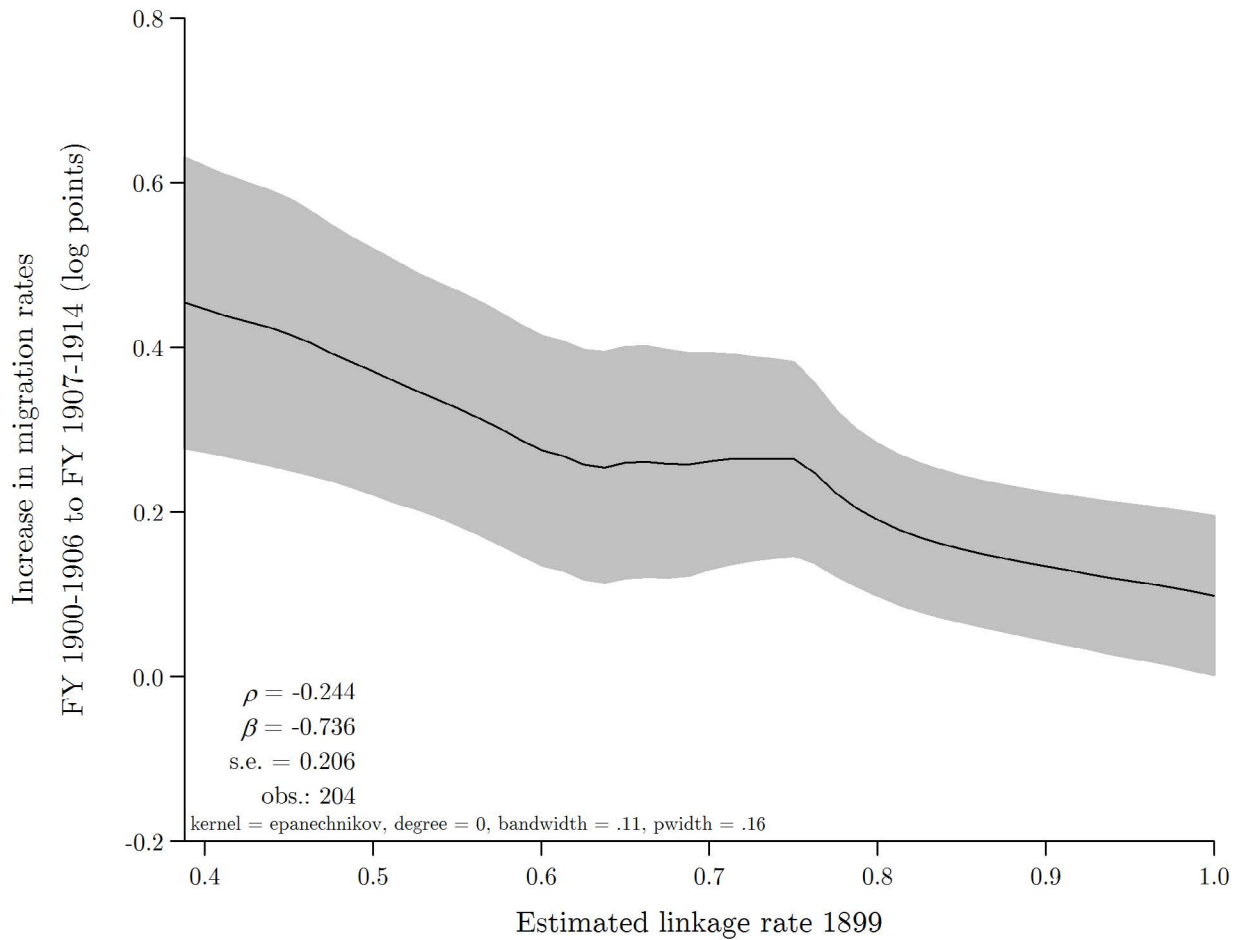
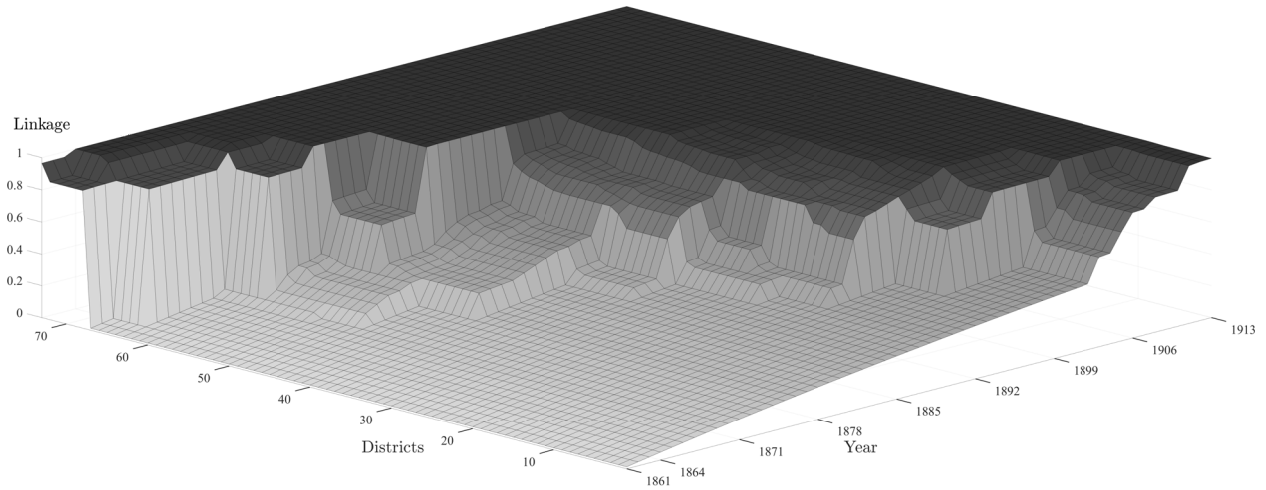


Figure 8: Linkage and acceleration in migration—Evidence of saturation

Notes: The figure plots a nonparametric regression of increase in migration on the rate of linkage. Each observation is a single district. The horizontal axis represents the estimates rate of linkage as of 1899. The vertical axis represents the increase in the average rate of migration of the cohorts aged 20–30 between FY 1900–1906 and FY 1907–1914. Two districts had zero migration recorded during the first period and were omitted from the plot.

(a) Poland



(b) New Russia

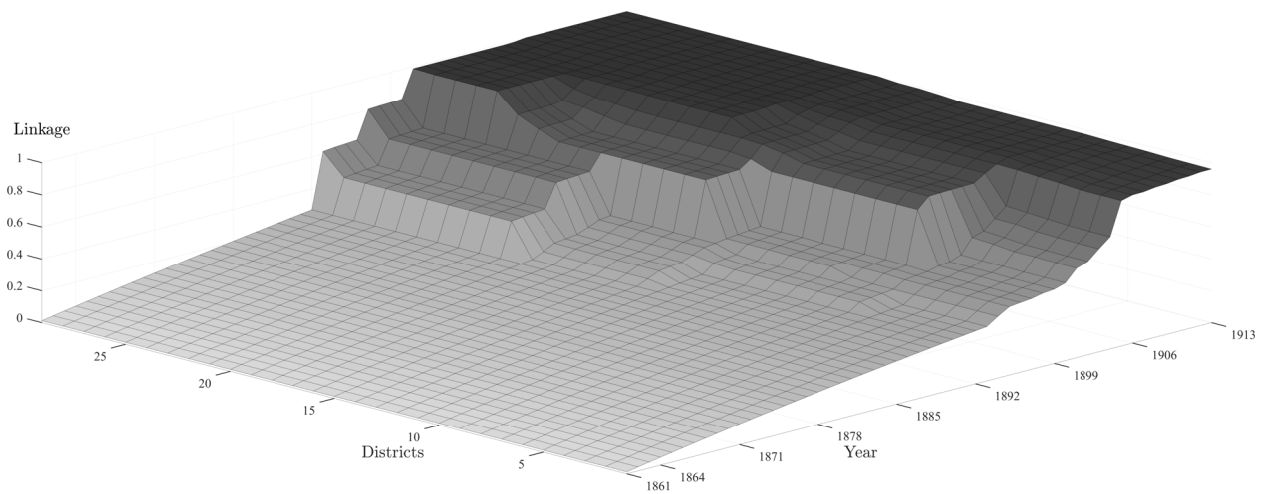


Figure 9: Diffusion of migration networks—regional estimated CDF

Notes: The figure represents the cumulative distribution function of the estimated linkage rates across districts for each year, in the regions of Poland and New Russia separately. For details see notes to figure 7.

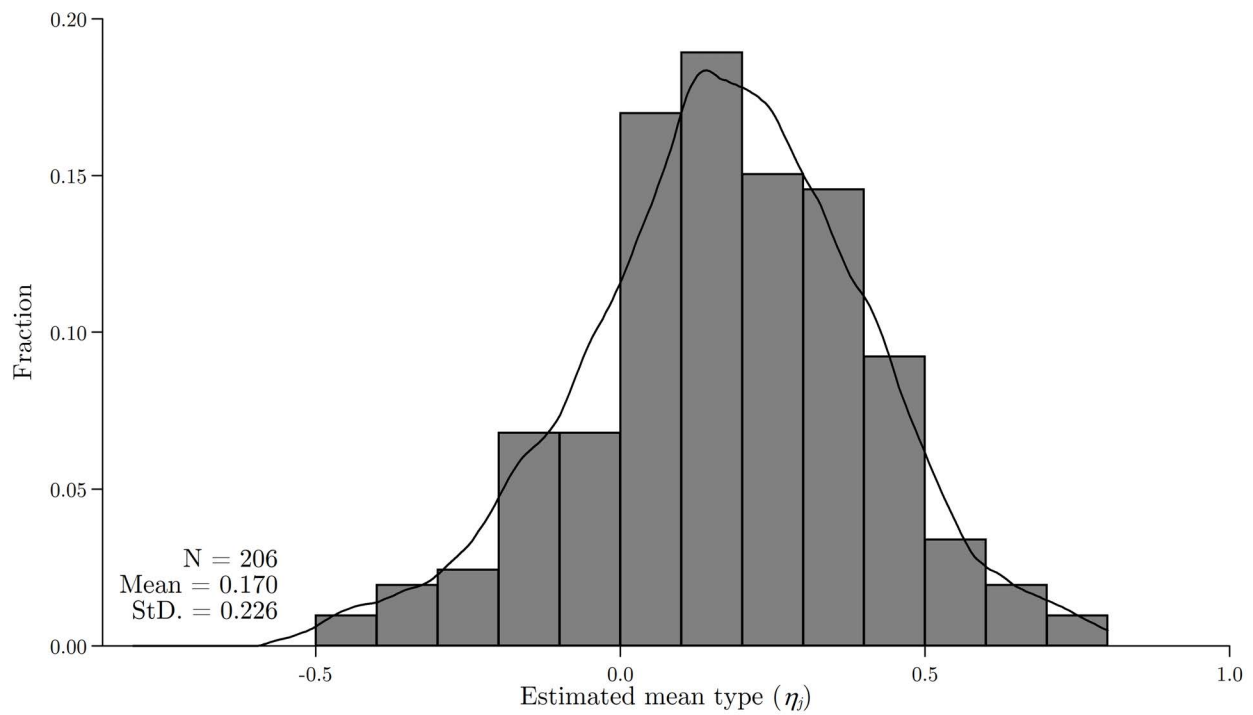


Figure 10: District fixed-effects

Notes: The figure presents a density plot of the distribution of district fixed-effects, the district specific mean type η_j . The estimates are from the benchmark structural model. The curve is the kernel density with Epanechnikov kernel and optimal half-width.

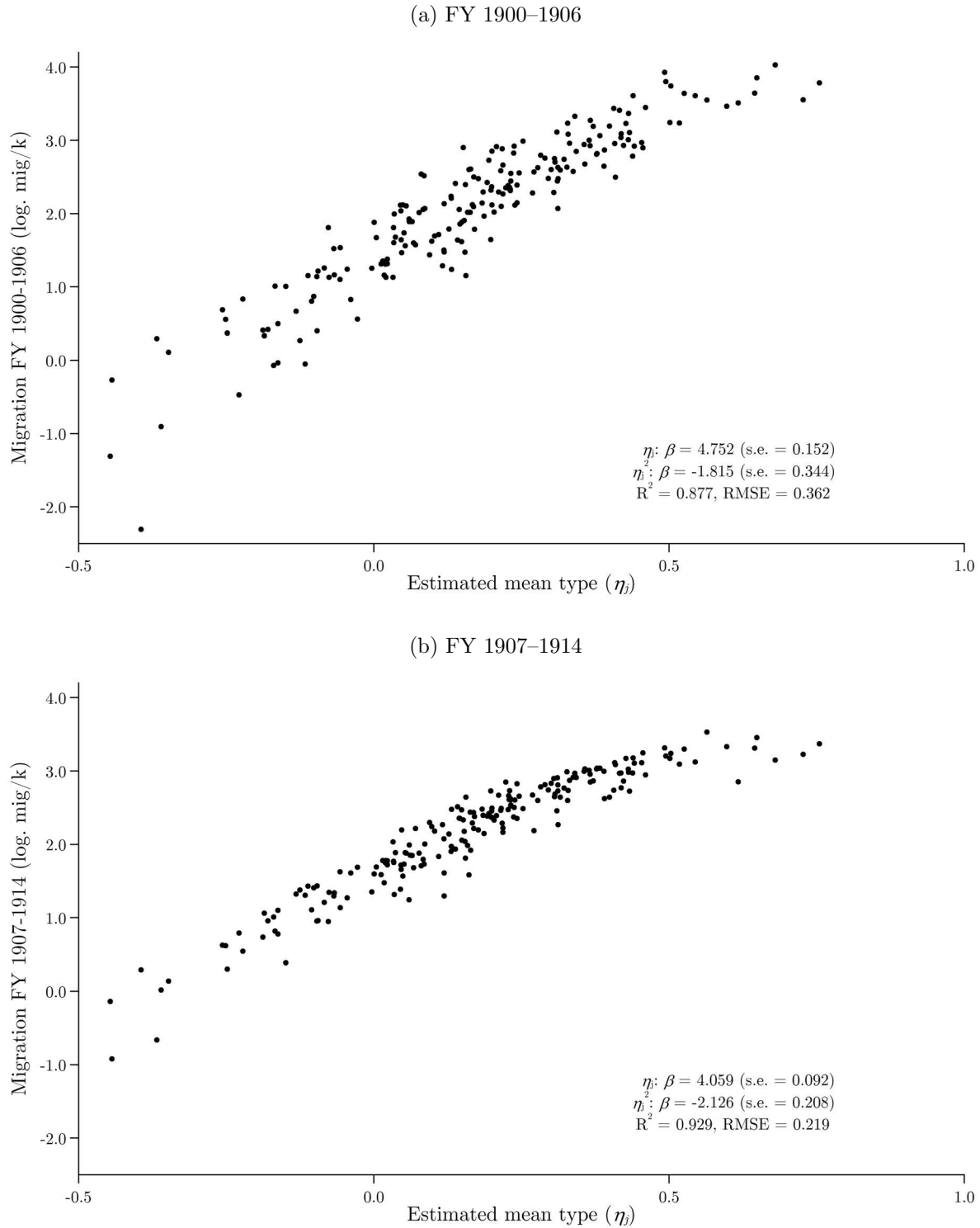


Figure 11: Migration and estimated district mean type

Notes: The plots present a scatter of the rates of migration plotted against the estimated district fixed-effects, according to the benchmark model. The horizontal has the district fixed-effects in utils. The vertical axis represents the log of the average yearly rates of migration per thousand of the cohorts aged 16–50 in FY 1900–1906 (upper plot) and FY 1907–1914 (lower plot). In both plots, the number of observations is 206. The reported statistics are from quadratic OLS regressions.

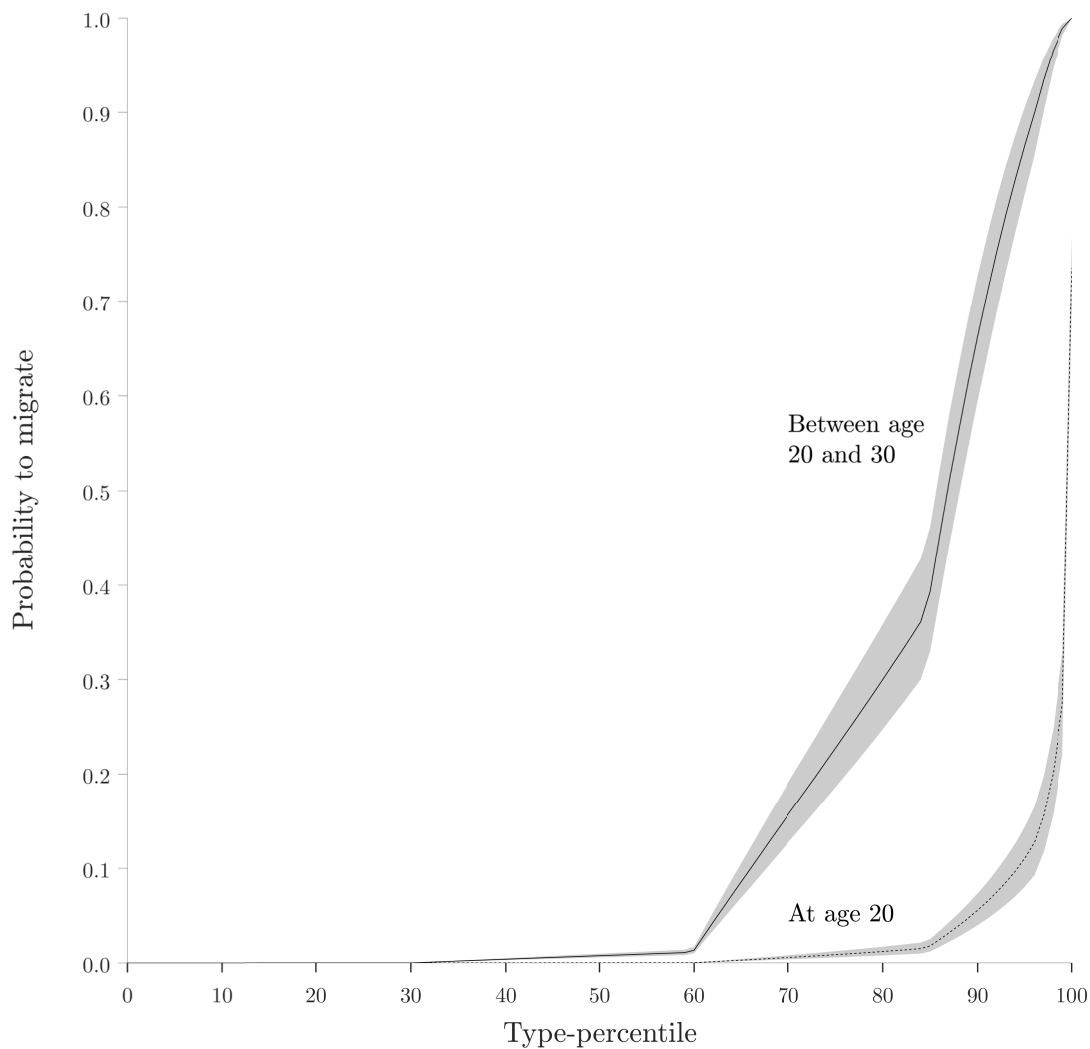


Figure 12: Probability to migrate by type

Notes: The dashed line represents the probability of an individual who is linked, and has not yet migrated, to migrate at age 20 at FY 1900, averaged over draws of BC shocks. The solid line represents the probability of the same individual to migrate at any time during the 11 years between age 20 and 30 (FY 1900-1910), averaged over draws of paths of BC shocks. The horizontal axis marks the type-percentile of the sample, according to the estimated Pale-wide distribution of types. The shaded areas represent 95 percent of the BC path draws (e.g., in 2.5 percent of the draws the probability to migrate is greater than the upper bound of the shaded area).

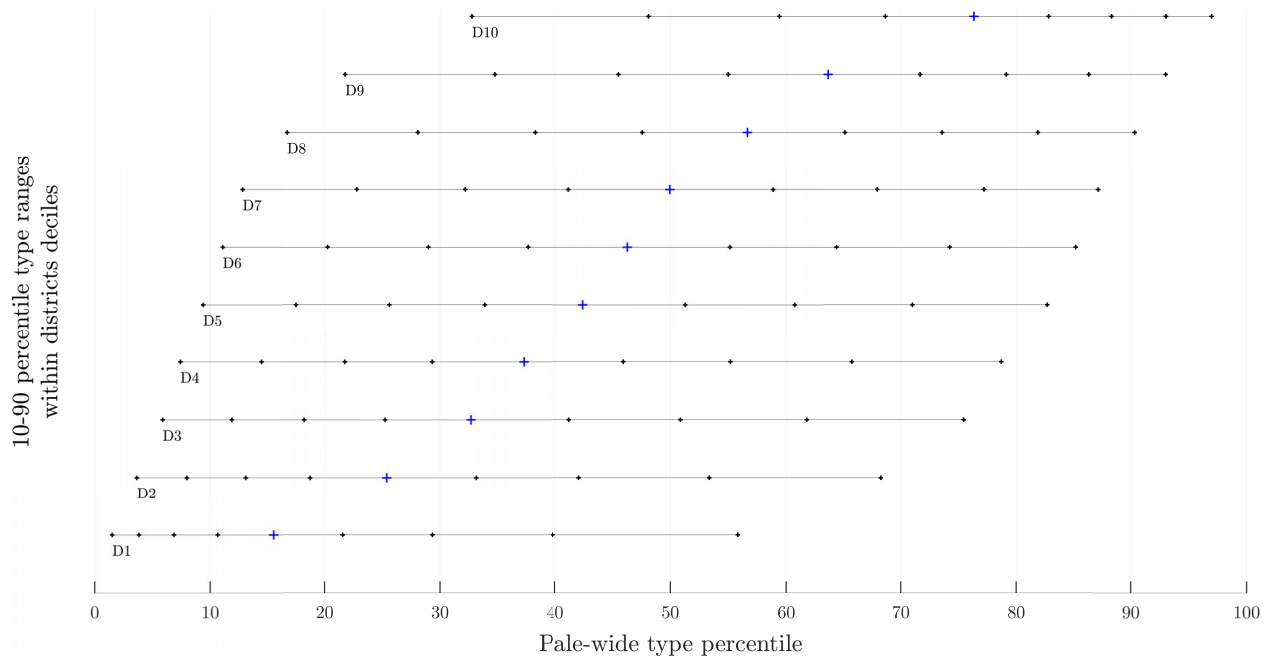


Figure 13: Distribution of types across and within district groups

Notes: The horizontal axis represents the Pale-wide type percentiles. Each of the ranges D1–D10 represent a decile of districts, ordered by their mean type (e.g., D10 is the group of districts with the top 10 percent of mean type). The ranges span the 10–90 percentiles of the within-group distribution of types. The markers along the ranges represent the within-group decile points, with the medians highlighted. The Pale-wide and within-decile group distributions of types are calculated based on the estimated districts mean types and the estimated types variation within districts, with the weight of each district being proportional to the size of the cohort that is 20 years old in 1900.

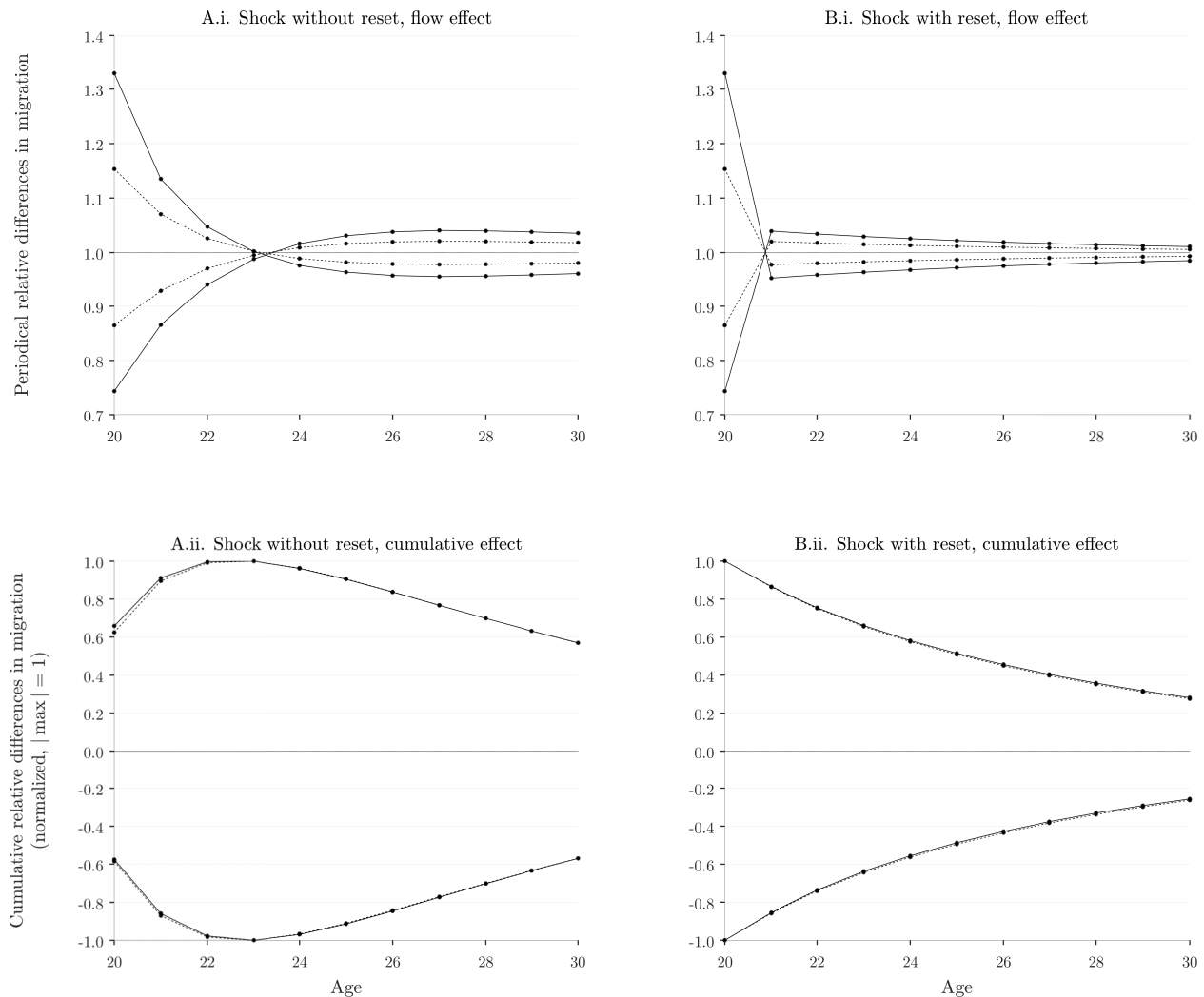


Figure 14: Distribution of types across and within district groups

Notes: The figures represent the effects of US business cycle shocks taking place in FY 1900, over the migration of a Pale-wide representative cohort of linked and present individuals aged 20 years old. Each figure depicts four shocks: plus or minus 1 or 2 standard deviations of deviation from income trend at FY 1900. In plots A.i. and A.ii. the income continues to evolve according to the estimated AR(1), with the first year shock persisting and decaying. The shocks in plots B.i. and B.ii. are reset, meaning that from the next year on the income evolves as if the true deviation from trend at FY 1900 was zero. Plots A.i. and B.i. report the ratio of migration under a given shock to migration under a no-shock scenario (i.e., FY 1900 income on trend). Plots A.ii. and B.ii. report the cumulative number of added migrants by each year, compared to a no-shock scenario, normalized as a share of their maximum absolute number over the path.. The curves represent averages over 260 simulated income paths (confidence intervals are suppressed for clarity).