Revisiting Italian emigration before the Great War: a test of the standard economic model

Author(s): PIER GIORGIO ARDENI and ANDREA GENTILI

Source: *European Review of Economic History*, NOVEMBER 2014, Vol. 18, No. 4 (NOVEMBER 2014), pp. 452-471

Published by: Oxford University Press

Stable URL: https://www.jstor.org/stable/43299740

*European Review of Economic History*, **18**, 452–471 
doi:10.1093/ereh/heu014 Advance Access publication September 17, 2014

# Revisiting Italian emigration before the Great War: a test of the standard economic model

PIER GIORGIO ARDENI* AND ANDREA GENTILI**
*Department of Economics, University of Bologna, Strada Maggiore 45, Bologna 40125, Italy, piergiorgio.ardeni@unibo.it*

Among the various statistical studies on migration before the Great War, Italy has received little attention, with a few notable exceptions. The standard economic approach explains emigration to any given country as driven by relative wages, relative employment rates, and the stock of previous emigrants to that country, the "network". This paper improves on earlier contributions by covering all Italian migration outflows between 1876 and 1915 to the most significant destination countries, taking all countries separately and simultaneously and adopting the most consistent and up-to-date econometric approaches. As it turns out, the standard model is only partially confirmed when accounting for heterogeneity of destinations, whereas other relevant hypotheses are not accepted.

## 1. Introduction

Between 1876 and 1915 more than fourteen million Italians left home, bound to either another European country or to the New World. For a country with a population of 27.3 million in 1871 and only 35.8 million in 1911, the outflow was equivalent to half the initial population stock, the largest ever migration outflow from Italy. The Great Italian Migration has naturally received considerable attention in the past, with an abundance of studies by historians, economic historians, and historical demographers who have delved into the *qualitative* and descriptive features of the phenomenon and their interpretation. Notable *quantitative* analyses have been those of Rosoli (1978) and Sori (1979), among others. And yet, *statistical* analyses, i.e., studies providing tests of economic hypotheses by means of statistical and econometric analyses—what is also called *empirical* literature—have been limited in number and scope in the case of Italy. In the last twenty years or so, quantitative economic analysis has been applied to international migration flows by adopting a standard "neoclassical" choice model and by testing its implications using econometric techniques.[1] This paper takes on that specific stream of studies by considering what we call the *standard economic model* and appropriately testing its founding assumptions. In particular, the paper aims at improving on earlier contributions *within that literature* to the understanding of the economic determinants of the Great Italian Migration, by checking the correctness of earlier results and the explanatory power of that framework of analysis.

The seminal work for that stream of literature can be traced back to Easterlin (1961b), who defined two main explanatory hypotheses for the Great Migration: the *Malthusian* hypothesis, whereby emigration originates from population pressure, and the standard neoclassic

*Pier Giorgio Ardeni is a Professor of Development Economics, Department of Economics, University of Bologna.
**Andrea Gentili is a Post-doc Researcher, Department of Economics, University of Bologna.
[1] See for all Hatton (1995) and Hatton and Williamson (1993, 1994, 1998).

*wage-differential* hypothesis, whereby emigration is simply due to a higher perspective salary in the receiving country. In the literature on migration, these explanations have also been known as the "push" and "pull-factor" hypotheses, respectively.

As we argue below, there are reasons to revisit Italian migration from the particular perspective of quantitative analysis. While the existing literature on Italian migration for the period before the First World War only considered a limited number of destination countries, we cover migration outflows to a larger number of countries. With such an enlarged pool of countries under observation, we try to replicate previous results, by using when possible the same data and sets of explanatory variables for the countries of destination. We show that the tests previously published are flawed and propose an alternative, correct estimation model. As it turns out, as the standard economic model appears to be only partially supported by the data, we believe that more convincing explanations—variables, data, and ways to treat them—are therefore needed.

Migration flows differed by country of destination, as Easterlin already pointed out in 1961 (Easterlin 1961b), citing Willcox: "As has frequently been observed, overseas emigration from Europe was characterized by wide swings, whose magnitude dwarfs the shorter term fluctuation associated with the business cycle".[2] As we show here, such *heterogeneity of destinations* has to be properly accounted for and would be totally overlooked by lumping all countries together (as for instance in Hatton and Williamson 1998).

This paper aims at improving on earlier empirical studies on the economic determinants of Italian migration before the Great War in three fundamental ways. First, we cover more destination countries than those previously considered and we treat them separately, thus accounting for differences in outflows by destination. Second, we show that previous studies are statistically flawed and we test the model correctly, using the most up-to-date methods on the same variables used therein, for the sake of comparison. Third, we properly test a number of other hypotheses that have been suggested in the literature, thus trying to disentangle what determinants actually drove Italian migration in the period under study.

As it turns out, the standard migration model—whereby outflows are proportional to differences in employment/activity rates, wage differentials, and the stock of previous emigrants—even with the most appropriate econometric techniques is only partially supported by the results under various specifications and only when the heterogeneity of destination countries is fully accounted for. There is no one model able to explain the Great Italian Migration outflows to economies as different and as changing as those to which Italians moved during the forty years before the First World War. This confirms that the *heterogeneity of destinations*, which is properly accounted for by the modeling approach used here by treating each receiving country separately, will have to be accompanied by a proper treatment of the *heterogeneity of origins*—by area, status, and living conditions of those who migrated—in order to fully explain the large migration flows that characterized Italy's initial development.[3]

## 2. Italian emigration in the recent literature

In line with the literature on European migration (briefly reviewed below in the Supplementary material, Appendix S1), Italian migration has been analyzed by a several economic-history

[2] See Easterlin (1961b, p. 341).

[3] We are currently exploring this hypothesis within a larger research project, of which this paper is a first building block.

studies, from Foerster (1919) to MacDonald (1958, 1963) and MacDonald and MacDonald (1964a,b) to Rosoli (1978) and Sori (1979). However, only a few have provided a convincing explanation of its overall economic and socio-demographic determinants in the vein of Hatton and Williamson's approach, most notably Faini and Venturini (1994), Moretti (1999), and Hatton and Williamson (1998).

At unification, in 1861, Italy was a relatively poor and mainly agricultural country (Faini and Venturini 1994) and yet its contribution to international migration was small. Out-migration began to pick up in earnest only in the 1870s, with a steadily increasing outflow reaching a peak in 1901 with more than 533,000 people leaving in one year, then again in the decade between 1905 and 1914 with more than 650,000 people annually.[4] The First World War led to a sudden decrease in the outflow and after that a dramatic change in the overall political and economic conditions brought to a permanent reversal of the migration pattern. Of those fourteen million Italians who in total migrated before the Great War, about six million people went to a country in Europe, while more than seven million went to the Americas (data are discussed at greater length in the Supplementary material, Appendix S2).

Most studies on Italian economic development and its industrial "take-off" have tended to rule out migration as a residual phenomenon—a "relief valve" for the excess supply of rural labor—adding to the explanation a mere pull-factor from the fast developing industrial economies of the United States and other European countries (like in Easterlin 1961a,b). However, something was missing in that analysis, as Italian migration cannot be explained only through Italy's late-comer industrialization status, as pointed out by Hatton and Williamson (1994). The overall determinants of the observed Italian migration outflow before the First World War have usually been ascribed by most historical studies to poor living conditions, excess supply of rural labor, underemployment, and lack of demand for labor in the lagging development of Italian industry (e.g., Vecchi 2011). These determinants were then translated as *explanatory* variables in a statistical regression model by the quantitative approach proposed by Hatton and Williamson and their followers: see, e.g., Hatton and Williamson (1993, 1994) and Del Boca and Venturini (2005) and, for Italy, Faini and Venturini (1994), Hatton and Williamson (1998), and Moretti (1999).[5]

In the case of Italy, as with other countries, migration models have incorporated three main hypotheses: First, population surpluses in Italy gave rise to excess in labor supply (both as underemployment and unemployment) bringing real wages down, thus making migration flows respond to real-wage differentials (Faini and Venturini 1994, building on Easterlin 1961b); second, more backward agriculture-based economies, like Italy's nineteenth-century, have lower incomes than developing industrial economies, thus making migration flow respond positively to income differentials, industrialization, and urbanization (Hatton and Williamson 1998, taking on Harris and Todaro 1970 and Hatton and Williamson 1994); and third labor-market conditions and the associated search for better jobs and higher wages are affected by the existence of "networks" and community links and are not necessarily only the result of individual choices—people tend to search for jobs where other family members or fellow countrymen live (Moretti 1999).

[4] Official comparable statistics are available before 1876 for only a limited number of destination countries, even if migration outflows, particularly to neighboring countries, were not irrelevant (as shown in Ferenczi and Willcox 1929, p. 811 and tables I–III, p. 817). In 1876 the *Annuario Statistico dell'Emigrazione* was started as an annual publication by the Italian General Directorate for Emigration, covering data for all destination countries. See Commissariato Generale dell'Emigrazione (1925).

[5] See also Gomellini and O'Grada (2011), although their focus is on migration at the regional level.

In all of these studies, however, the "standard" model does not seem to explain migration outflows satisfyingly when applied to the Italian case, at least in purely statistical terms. In particular, it appears that wage differentials and relative employment rates do not explain the large population movements that occurred between Italy and the rest of the world during the four decades before 1915. Hatton and Williamson (1998) asked "to what extent can Italian emigration be explained by the same economic and demographic forces identified" for other European countries, only to conclude that in the case of Italy "the determinants of Italian emigration are much too complex to be isolated by [bivariate correlation] analysis" and that "that conventional wisdom is often badly bruised when exposed to multivariate analysis" (p. 96). Their analysis, in any case, supported the standard migration model when applied to the overall aggregate migration flow from Italy to "the rest of the world".[6] On the other hand, Faini and Venturini (1994) examined migration to France, Germany, and the United States only, while Moretti (1999) focused on four overseas destination countries only—United States, Canada, Argentina, and Brazil. All of these works did not account for any type of migration costs.[7]

Faini and Venturini (1994) analyzed the importance both of demographic changes and of structural changes in the economy, also focusing on the so-called *poverty trap*,[8] while Moretti (1999) stressed the importance of the so-called *network effect*—the importance of "friends and relatives". The case of Italian emigration was termed by Moretti as "puzzling". Given the large gap in expected returns (both in terms of present and future values) out-migration should have been much larger at the beginning of the period, only to slowly decrease as time went on. However, observed wage differentials and employment rates seem to tell a different story (see Figure 1, with four countries as an example). As the four graphs show, there is no apparent correlation between Italian migration flows to the United States, France, Germany, and Brazil and the relative wage ratios (in the case of United States and Germany) or the employment rates (in the case of France and Brazil).

Both Faini and Venturini (1994) and Moretti (1999) pointed out that a large part of Italian migration took place only beginning in the 1880s, when Italian industrial development started to pick up. Specifically, Moretti emphasized how—purportedly—there was no response to wage differentials in Italian emigration to the Americas. As wage differentials started to decrease, so migration increased. Italy's emigration rose from five per thousand in 1876 to twenty-five per thousand in 1913[9] while wage gaps decreased or remained stable. Even relative employment levels (or rates), as suggested by Gould (1979), do not seem to explain migration outflows during that period. As pointed out in Faini and Venturini (1994), the Italian employment growth rate was below the main destination countries' until 1900, but it was in line with or above them after 1900, when migration flows started to pick up, as in response to a "poverty trap".[10]

[6] Hatton and Williamson (1998) considered the aggregated migration flows from Italy to *all* destination countries taken together (the "rest of the world") from 1876 to 1915, using the data taken from the *Annuario Statistico dell'Emigrazione* referred to above. As for the main economic variables in the receiving countries, they considered some averages of the main variables *for five countries only:* France, Germany, UK, Argentina, and the United States.

[7] The only other study on Italian migration covering the same period—Gomellini and O'Grada (2011)—focused on Italian regional disparities and their differences in contributing to out-migration, without analyzing migration by country of destination.

[8] As in Faini and Venturini (1993, 2010).

[9] The number of emigrants is usually reported in terms of thousand residents at the start of the reference period.

[10] We should also mention Hatton and Williamson (1998) discussion (already taken up by Gould, 1979) on the "big surge" in Italian migration around the turn of the century. In their opinion, that surge was spurious (and exaggerated), due to the change in the administrative data collection procedures. This would be confirmed by the difference

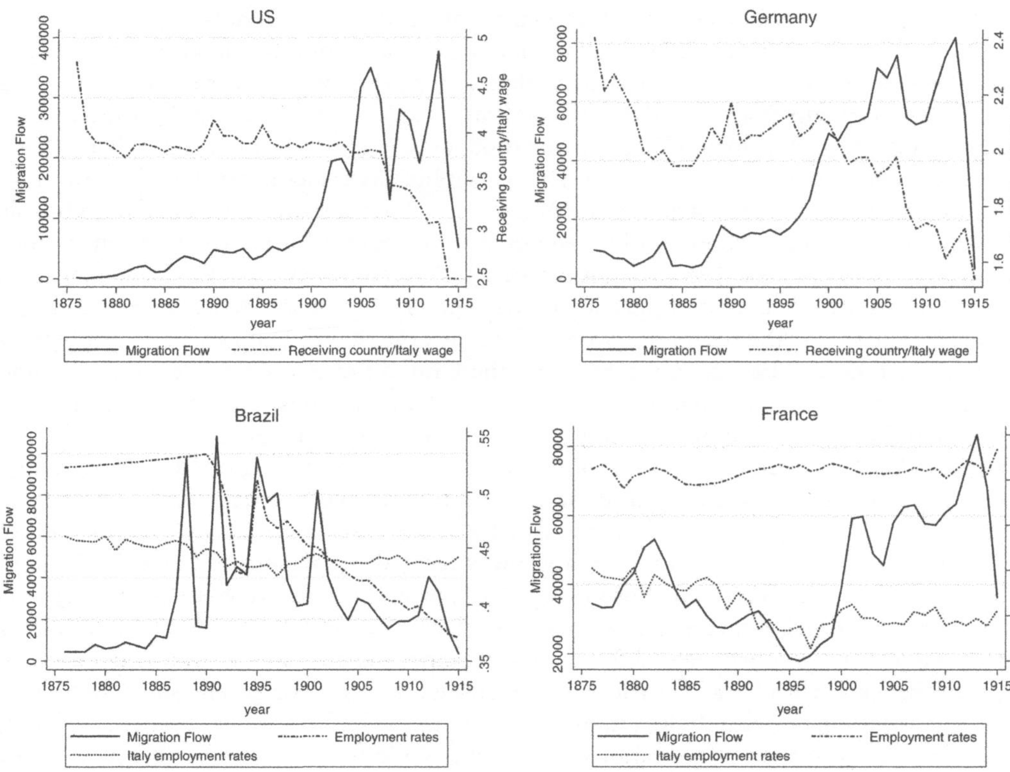


Figure 1. *Italian migration to United States, Germany, Brazil, and France.*

We will see below how this explanation fares when the statistical model is put to test. As it will turn out, all of the studies above provide an explanatory framework that is not supported by the evidence and is possibly statistically flawed. After Hatton and Williamson and the others, it was thought that Italian migration responses to the basic economic determinants had been no different from all other main countries that were characterized by large migration flows in their early stages of development. And yet, as we will see, those approaches fail to pass the formal testing and their underlying explanations are not fully supported by the data.

## 3. The "standard" model and the data

In light of the data presented above and the literature discussed before, two questions emerge. First, are the quantitative explanations provided for the other main European countries that have witnessed relevant migration outflows statistically valid in the case of Italy? Second,

between emigration data from Italy to the United States, which are systematically lower before 1901 and systematically higher after 1901. This, in any case, does not imply that the *net migration* rates should be decreasing, as shown in their book (on p. 97), as return migration statistics are quite questionable for that whole period.

were there underlying determinants (i.e., statistical explanations) that made Italy differ from other countries?

After Gould's (1979) criticism of the empirical evidence of that time, Hatton and Williamson—with their various contributions in the 1990s—refined the modeling approach and provided new highlights by the use of new data, thus defining what we may now call the *standard quantitative migration model.* As we mentioned above, using that modeling approach, Italy's migration has then been studied, most notably, by Faini and Venturini (1994), Hatton and Williamson (1998) and Moretti (1999). As also Hatton and Williamson (1998) acknowledged, the Italian case is all the more important if we consider that millions of people migrated from Italy over the period and Italy represents in many respects a relevant "emigration story". And yet, those few studies provided results that were somehow unsatisfactory, questions remained somehow unexplained, and several issues were left unanswered. The question is then whether Italy's emigration cannot be explained by the standard model because of poor model specification, poor empirical evidence, or poor data.

First, some studies focused on a *limited number* of destination countries (Faini and Venturini (1994; Moretti 1999), covering only part of the total migration flows. Also, migration outflows to specific countries appear to be quite uncorrelated to wage rate differentials as well as to employment level differentials.[11] Moreover, migration costs as measured by time trends do not meaningfully explain overseas migration, as opposed to migration to Europe (while distance or some more precise measure would better account for the choice of destination). Finally, there is also an issue of poor model specification and improper treatment of the data as the empirical literature on Italian migration, even after Hatton and Williamson (1993, 1994, 1998), does not seem to correctly account for the time-series properties of the data, like heteroskedasticity, auto-correlation, non-stationarity, and co-integration (these issues are explored in the Supplementary material, Appendix S3).

We model Italian migration for the 1876–1914 period by adopting the theoretical framework, which we term the *standard economic model* and whose closest reference is Hatton (1995)

$$m_{i,t} = \alpha_0 + \alpha_1 \ln\left(\frac{w_{i,t}}{w_{h,t}}\right) + \alpha_2 \ln(e_{i,t}) + \alpha_3 \ln(e_{h,t}) + \alpha_4 S_{i,t} + \alpha_5 d_{i,t}, \qquad (1)$$

where $m_{i,t}$ is the *migration flow* at time $t$ from Italy to country $i$—or, alternatively, the *migration rate* obtained by dividing the flow by the population of Italy (so as to have a rate expressed in thousands); $\ln(w_{i,t}, w_{h,t})$ is the logarithm of the ratio between the wage rate in country $i$ and the Italian wage rate,[12] $\ln(e_{i,t})$ is the logarithm of the aggregate level of employment in country $i$, $\ln(e_{h,t})$ is the logarithm of the aggregate level of employment in the home country (Italy), $S_{i,t}$ is the stock of Italian migrants to country $i$, calculated as the population in country $i$ of Italian origin until time $t-1$ (the previous year) plus the number of migrants of the previous year, thus measuring the importance of the "network effect",[13] and $d_{i,t}$ is a cost variable, measuring the global "cost" of migrating from Italy to country $i$ at time $t$, which is proportional to distance.[14] All coefficients should be positive, with the exclusion of the ones for

[11] Italian migration outflows show quite some variability from year to year (a common feature to many other countries), which implies that short-run determinants and cyclical effects must have been important. Even so, long-run trends are quite clear.

[12] As described above, data are from Williamson (1995).

[13] For both employment ratios and network variables, the data are as described in the Supplementary material, Appendix S2.

[14] Strictly speaking, the cost variable for not included in the original Hatton (1995) model.

Italian employment and for cost: as wage ratios, employment in the receiving country and network effects increase, migration flows should also increase.

The *empirical* model as specified in equation (1) and tested below is actually different from both Hatton (1995) and Hatton and Williamson (1998)—which focused on Italian migration—in that here we consider the migration flow from Italy to each country *separately*, so as to account for the compound nature of the migration choice. While Hatton (1995) empirical specification estimated the composite effect of wages and employment levels of three destination countries on British migration flows *all at once*, and Hatton and Williamson (1998) considered the effect on Italian emigration of the economic variables of five countries *at once*, we are able to estimate such effects separately and for each destination, by estimating one equation for each country.[15] This way we are able to account not only for the fundamental binary choice between staying or migrating, but also for the choice where to migrate.[16] This, in our opinion, is also the proper test of the theoretical model proposed by Hatton (1995).

We used the data on migration flows to eleven countries or groups of countries, covering sixteen destination countries and accounting for about 85–90 percent of total migration over the period.[17] As for the "explanatory" variables, we used Williamson's wage rates; employment rates in Italy and in the receiving countries or combinations of them in the case of country groups; "network" variables and distance variables. Wage rates were measured by purchasing-power-adjusted minimum wages in the industrial sector for low-skill jobs.[18] Employment rates were computed, for each country considered, by interpolating census data on employment status (or on economically active population in the case of Brazil, as in Mitchell 1993).[19] As for the "network" variables, as measured by the stock of previous migrants, we estimated the stock of population of Italian origin in all of the receiving countries included.[20] Finally, distance was proxied with the distance between the capital of country $i$ and Rome (in thousands of km) divided by a linear time trend to take into account the decreasing cost of migration due to improvement in technology and infrastructures.[21] In the case of country groups, for all the above variables, we used weighted averages, with country populations as weights.

[15] In Hatton (1995), "The wage and employment rates for the United States, Canada, and Australia were combined to form an average for overseas destinations" (p. 411). The dependent variable measured total UK migration to those three countries. In Hatton and Williamson (1998), "the variables representing foreign conditions are weighted averages for five major receiving countries: France, Germany, Argentina, Brazil, and the United States" (p. 103).

[16] This also helps accounting for the "spread" of migration flows to an increasing number of countries shown by Italy over the period (Hatton and Williamson, 1998).

[17] The eleven country groups are Belgium, Netherlands and Luxembourg (treated as one economy, Benelux), France, Germany, Great Britain, Spain and Portugal (treated as one economy, Iberia), Denmark, Norway and Sweden (treated as one economy, Scandinavia), Argentina, Australia, Brazil, Canada and the United States. We have excluded from our empirical analysis Switzerland, Austria, North Africa, and the Middle East. For those receiving countries, in fact, we do not have a set of the same relevant economic variables that are available for other countries, including Williamson's (1995) wage rates. As for Austria–Hungary, see footnote 21.

[18] Wage rates were constructed by Williamson using wages for unskilled workers in the construction sector and by interpolating over time when no information was available, using census "pillars", like in the case of Italy. Employment levels, too, for most countries come from estimates constructed by using decennial census data as "pillars". This, as we will see, might be the reason for some persistence or trend-stationarity in the data.

[19] Interpolations were done using alternatively three methods: linear interpolation over time, the variation of GDP using Maddison's GDP data, and the cumulated GDP variation on Maddison's data.

[20] Based, alternatively, on census data of the receiving countries on the number of people born in Italy; the Italian census of Italians living abroad and Ravenstein (1889) data.

[21] Thus, the distance variable can also be seen as a *cost of transportation* variable. We also tried distance over railways km (in Italy), distance over international boat arrivals, and similar information without obtaining significant differences.

The data we have used are based on a large database that we have constructed by collecting the data on migration *at the sources*—the original statistical yearbooks, the censuses—and that we have cross-checked for consistency from the provincial to the regional and national level and across all other variables (see the Supplementary material, Appendix S2).[22] We have covered an unprecedented number of countries—unlike any other study after Hatton and Williamson's—and we have constructed "network" variables for a large set of receiving countries. The standard migration model aims at providing a *universal* explanation to all migration flows, abstracting from any country specificity. Yet, the migrant's choice of a destination country depends on both the short- and the long-term country-specific dynamics as well as on how much one can switch from one country to the other. We believe that it is of utmost importance to keep destination countries separate and that all of them be included in the analysis (or at least the maximum number of them possible). Lumping all countries together or including only a few would not allow to properly account for such "substitution effects" across countries.

Before estimating equation (1), we looked at the time-series properties of the variables (see the Supplementary material, Appendix S3). As it turns out, the non-stationarity and co-integration features of the data call for an appropriate treatment of the series, accordingly.[23] Different countries of destination have variable with different dynamics, which calls for a separate and simultaneous treatment of the series able to account for different, albeit possibly converging, short- and long-run dynamics. Variables showing different stochastic properties across countries confirm the existence of a heterogeneity at destination that can only be accounted for by the separate and simultaneous inclusion of all countries.

## 4. Putting the model to test: the results

We present here two exercises. First, we test the model by estimating *four* different versions of equation (1). Second, we test a number of hypotheses that have been suggested in the literature as the main drivers of migration flows.

The *first testable version* of the model is adapted from Hatton and Williamson (1998) application to Italian data and is the *simplest version* of the "standard" model. Equation (1) can be written as:

$$m_{i,t} = \alpha_0 + \alpha_1 \ln\left(\frac{w_{i,t}}{w_{h,t}}\right) + \alpha_2 \ln(e_{i,t}) + \alpha_3 \ln(e_{h,t}) + \alpha_4 S_{i,t-1} + \alpha_5 d_{i,t} + \alpha_6 m_{i,t-1} + \alpha_7 m_{i,t-2} + \alpha_8 n_{h,t} + \alpha_9 D0113. \quad (2)$$

This is a variation of Hatton and Williamson's (1994) estimation equation for Italy and deserves a few comments.[24] In the first place, this is the *panel-regression version* of that specification, as we estimate an eleven-equation model for eleven different destination countries, thus accounting for the specificity and the heterogeneity of migration to each country. Thus,

[22] Other studies, like Ferenczi and Willcox (1929), have published similar data on migration, with some "blanks" here and there. Rosoli (1978) also reported migration data for the same period, albeit in a summarized form.

[23] For instance, aggregating across countries series with different and nonhomogeneous stochastic properties would be incorrect.

[24] The exact definition of the migrant stock should not be misleading. Here by "stock at time $t-1$" we mean "*at the end of* $t-1$", while Hatton (1995), for instance, uses the stock at time $t$ as the stock *at the beginning of time t*. The two definitions are, for that matter, identical.

all variables are country-specific and are not aggregated with dubious weighing systems.[25] Second, in our specification, we keep the effects of the two "activity" variables—that in country *i* and that at home—separate and simultaneous, as we see no reasons that justify a different expectation mechanism for employment perspectives at home and abroad. Third, we define the "log rate of natural population increase" similarly to Hatton and Williamson (1998) and Faini and Venturini (1994) as the "increase in population 15 years before".[26] Fourth, the "cost" variable thus defined takes into account distances among countries, while at the same time accounting for the time-trend effect (thus replicating Hatton and Williamson 1998). Finally, similarly to Hatton and Williamson (1998), we introduce a dummy variable to account for the potentially "spurious" big-surge effect in the data from 1901 to 1913, which we term *D0113*.

Results of estimation of equation (2) are reported in the first column of table 1, and they show that the model is not supported by the data.[27] The only significant variables are the network variable, the lagged dependent variable, and the dummy variable. Equation (2) was estimated using a standard *homogeneous panel-regression* model: that model assumes that all coefficients are equal across equations, which is obviously quite a strong assumption contradicting our tenet that the high heterogeneity across countries is bound to have different effects across all variables over time and must therefore be taken into account.

We have thus tried a *second version* of equation (1) above, adapted from Hatton (1995)

$$\begin{aligned} m_{i,t} = \alpha_0 + \alpha_1 \ln\left(\frac{w_{i,t-1}}{w_{h,t-1}}\right) + \alpha_2 \ln(e_{i,t-1}) + \alpha_3 \ln(e_{h,t-1}) + \alpha_4 S_{i,t-1} \\ + \alpha_5 d_{i,t} + \alpha_6 m_{i,t-1} + \alpha_7 \Delta\left(\frac{w_{i,t}}{w_{h,t}}\right) + \alpha_8 \Delta \ln(e_{i,t}) + \alpha_9 \Delta \ln(e_{h,t}), \end{aligned} \tag{3}$$

where differently from equation (2), the wage and activity variables are lagged one period, the dependent variable has one lag only, and the first differences of the wage and activity variables are also introduced. The justification for having the first differences of the three variables given in Hatton and Williamson (1998) was that the theoretical model can be rewritten "in the form of a simple first-order correction mechanism" and then estimated.[28]

Results for this version are presented in the second column of the table 1. Once again, it appears that the model is not supported by the data, as the only significant variables are the lagged dependent one and the network variable. Again, equation (3) was estimated using a standard *homogeneous panel-regression model*, which is not the appropriate estimation strategy for the case at hand, as we argued above.

[25] In our opinion, it would be difficult to define a "proper" weighing criterion in this case, as using the share of migrants over the total number of migrants appears to generate endogeneity problems, while any other criterion would be debatable.

[26] Both Hatton and Williamson (1998) and Faini and Venturini (1994) used the "population growth rate 20 years before". We used a time lag of 15 years not only because our series begin in 1876 and Italy was unified in 1861 (15 years before) but also because, in order to measure population "pressure", it makes more sense to use an age limit that coincides with the working age limit and with the minimum age required to obtain a passport at the time.

[27] Notice that the all specifications in tables 2 and 3 were estimated for the period 1878–1913 as the introduction of two lags "cuts off" the beginning of the sample period, while as for the end year, 1914, some of the countries in the sample were already at war.

[28] But then, their theoretical specification had a first-differenced migration flow as the dependent variable, while the estimates presented in their paper had the migration flow variable in levels.

Table 1. *Homogeneous panel approach–Hatton and Williamson revisited*

| Migration rate per thousands 1878–1913 | | | | | |
|---|---|---|---|---|---|
| $m_t$ | Hatton (1995) | $m_t$ | Hatton and Williamson (1994) | $\Delta m_t$ | Differences and levels |
| Constant | 0.991 (1.19) | Constant | 0.230 (0.21) | Constant | −0.234 (−0.87) |
| $\mathrm{Ln}(w_f/w_h)_t$ | 0.268 (1.30) | $\mathrm{Ln}(w_f/w_h)_{t-1}$ | 0.229 (1.23) | $\mathrm{Ln}(w_f/w_h)_{t-1}$ | 0.017 (0.53) |
| $\mathrm{Ln}(e_f)_t$ | 0.106 (0.45) | $\mathrm{Ln}(e_f)_{t-1}$ | 0.027 (0.12) | $\mathrm{Ln}(e_f)_{t-1}$ | −0.048 (−0.87) |
| $\mathrm{Ln}(e_h)_t$ | 1.171 (1.34) | $\mathrm{Ln}(e_h)_{t-1}$ | 0.349 (0.28) | $\mathrm{Ln}(e_h)_{t-1}$ | −0.239 (−0.78) |
| $\Delta pop_{t-15}$ | −10.877 (−1.46) | | | | |
| $M_{t-1}$ | 0.653*** (13.93) | $M_{t-1}$ | 0.785*** (17.48) | $M_{t-1}$ | −0.028** (−1.91) |
| $M_{t-2}$ | 0.175** (2.40) | | | | |
| $\mathrm{Network}_{t-1}$ | 0.0007** (1.99) | $\mathrm{Network}_{t-1}$ | 0.001*** (2.60) | $\mathrm{Network}_{t-2}$ | 0.0001* (1.71) |
| 1901–1913 | 0.123* (0.081) | | | | |
| $C_t$ | −0.002 (−1.52) | $C_t$ | −0.002 (−1.55) | $C_{t-1}$ | −0.007 (−0.30) |
| | | $\Delta\mathrm{Ln}(w_f/w_h)_t$ | 0.920 (1.42) | $\Delta\mathrm{Ln}(w_f/w_h)_t$ | 0.077 (1.48) |
| | | $\Delta\ln(e_f)_t$ | 0.711 (0.94) | $\Delta\ln(e_f)_t$ | 0.152 (0.85) |
| | | $\Delta\ln(e_h)_t$ | 1.565 (1.28) | $\Delta\ln(e_h)_t$ | −0.364 (−0.58) |
| | | | | $\Delta C_t$ | −0.142 (−0.28) |
| | | | | $\Delta\mathrm{Network}_{t-1}$ | 0.055*** (5.40) |
| Obs | 374 | | 374 | | 374 |
| Groups | 11 | | 11 | | 11 |
| $R^2$ | 0.83 | | 0.82 | | 0.83 |

*, **, and ***are 10, 5 and 1 percent significance level, respectively.

Given that most of the variables appear to be non-stationary and co-integrated (as discussed in the Supplementary material, Appendix S3), as it was also the case in Hatton (1995)[29], we thus estimated a *third version* of model (1) above, which can be interpreted as the "appropriate" estimating equation of Hatton (1995) model[30]

$$\Delta m_{i,t} = \alpha_0 + \alpha_1 \ln\left(\frac{w_{i,t-1}}{w_{h,t-1}}\right) + \alpha_2 \ln(e_{i,t-1}) + \alpha_3 \ln(e_{h,t-1}) + \alpha_4 S_{i,t-2} + \alpha_5 d_{i,t} + \alpha_6 m_{i,t-1} + \alpha_7 \Delta\left(\frac{w_{i,t}}{w_{h,t}}\right) + \alpha_8 \Delta \ln(e_{i,t}) + \alpha_9 \Delta \ln(e_{h,t}) + \alpha_{10} \Delta S_{i,t-1} + \alpha_{11} \Delta d_{i,t}, \tag{4}$$

where all variables appear both in levels and first differences and the dependent variable is the first difference of the migration rate. Results for this version are reported in the third column of table 1. Once again, it appears that the model is not supported, as the only significant variables are the lagged dependent one and the network variable, both in levels and in first differences. In the case of equation (4) the same caveats hold, as this is a homogeneous panel-regression model, which is not appropriate for the case at hand.

Given the results above, we consider that the proper estimation strategy would be that of a *panel-regression model*. A *homogeneous* panel model, however, would not be appropriate in this case, as we have clearly seen that migration to different countries followed different patterns

[29] Even though, in that case, it was a one-equation co-integration regression.
[30] See equation (10) in Hatton (1995, p. 410).

Table 2. *Heterogeneous panel approach–Basic model*

| Migration rate 1878–1913 Emp on GDP | | | |
|---|---|---|---|
| Pooled mean group estimation | | | |
| Convergence coefficient | | | |
| $M_{t-1}$ | −0.046* (−1.79) | | |
| Long-run coefficients | | Short-run coefficients | |
| $\mathrm{Ln}(w_f/w_h)_{t-1}$ | 6.565** (2.38) | $\Delta\mathrm{Ln}(w_f/w_h)$ | 0.856 (1.25) |
| $\mathrm{Ln}(e_f)_{t-1}$ | 12.012* (1.65) | $\Delta\mathrm{Ln}(e_f)$ | 1.417* (1.77) |
| $\mathrm{Ln}(e_h)_{t-1}$ | 19.913** (2.19) | $\Delta\mathrm{Ln}(e_h)$ | 0.032 (0.08) |
| $\mathrm{Network}_{t-2}$ | 0.006* (1.89) | $\Delta$Network | 0.038** (2.04) |
| $C_{t-1}$ | −0.289 (−0.32) | $\Delta C$ | −3.232 (−1.19) |
| | | Constant | 18.851 (2.00) |
| Number of areas | | | 11 |
| Observation | | | 385 |
| Log likelihood | | | 1037.55 |

*, **, and ***are 10, 5, and 1 percent significance level respectively. Dependent variable $\Delta m_t$.

over time, reacting with specific coefficients of adjustment to the given set of variables, and should thus be treated separately, albeit simultaneously. We have therefore adopted the *Pooled Mean Group* (PMG) approach by Pesaran *et al.* (1999) for *heterogeneous* panels, as the PMG allows us to reconcile some of the migration model assumptions with the needs of panel-data regressions. In particular, it lets short-term adjustments and convergence speeds to vary across destination countries, and, by considering equation (1) as a long-run equilibrium relationship, it imposes cross-country homogeneity restrictions on the long-run coefficients only. Thus, in this specification, we have common long-run coefficients across all receiving countries, as migration determinants should eventually converge to the same values in the long run (the "steady state"). Conversely, we allow the speed of convergence to the steady state to differ across countries. Assuming equal short-run coefficients across countries (equations) would be consistent with the standard migration model only if migration determinants were the same and equal across countries in any given time (the homogeneous models estimated above). The *error-correction* version of model (1) is thus

$$\Delta m_{i,t} = \alpha_0 - \theta\left( m_{i,t-1} - \beta_1 \ln\left(\frac{w_{i,t-1}}{w_{h,t-1}}\right) - \beta_2 \ln(e_{i,t-1}) - \beta_3 \ln(e_{h,t-1}) - \beta_4 S_{i,t-2} - \beta_5 d_{i,t} \right) + \alpha_7 \Delta\left(\frac{w_{i,t}}{w_{h,t}}\right) + \alpha_8 \Delta\ln(e_{i,t}) + \alpha_9 \Delta\ln(e_{h,t}) + \alpha_{10}\Delta S_{i,t-1} + \alpha_{11}. \tag{5}$$

We estimated equation (5) using the panel-regression PMG estimator.[31] Column 1 in table 2 reports the estimation result for the long-run coefficients—the vector error-correction accounting for co-integration, i.e., the component in brackets—and their average (across equations) short-run counterparts (coefficients are reported on the row, $z$-statistics are in parentheses). As the table shows, the estimated panel-regression model derived from the standard migration model finally seems to show a better performance than the previous specification versions. Most variables are significant and their coefficients generally have the expected signs, with one

[31] Using the XTPMG.ADO procedure for Stata as proposed by Balckburne and Frank (2007).

relevant exception: the home activity level as measured by employment. Overall, wage ratios, foreign activity levels and migrant networks seem to have a positive effect on Italian migration in the long run, as expected, together with domestic employment levels, which is not in line with the model. In the short run, there are notable differences, although the coefficients should be taken with care, as they average out potential differences in the speed of adjustment across countries. The estimated "speed of convergence", the estimated $\alpha_1$ coefficient is equal to 4.63, which implies a very slow convergence to the long-run equilibrium level.

Even though the model seems to be partially supported, the sign and value of the "home activity" coefficient appear to be mystifying. We have therefore explored this issue in greater detail and looked at alternative specifications of the activity variable, to see whether this is an issue of model misspecification or "wrong" data. What we call home activity—measured by the level of employment—is actually a proxy for the *demand for labor*, which, as we know, is not necessarily shown in employment data.[32] In the absence of a good proxy for the demand for labor, it is customary to rely on variables describing the level of employment. Also, as no annual employment level data are available for most countries for the period, such variables are generally constructed by interpolating employment level figures usually derived from decennial census data, using some variable representing the level of economic activity. Yet, as it turns out, interpolation methods are not neutral. While a simple linear trend would introduce an obviously spurious smooth behavior over time, other interpolation methods may give rise to spurious volatility. The issue somehow leads to what we expect such "activity" variable to express. If we interpolate employment levels with a variable representing the short-run behavior of the business cycle or gross domestic product (GDP), then it will show the typical volatility of short-run variations in the level of economic activity. If we interpolate employment levels with the cumulative stock of GDP variations, for instance, then it will reflect the long-run growth of the economy.[33]

In conclusion, the model appears to be extremely sensitive to how the variables are treated and how the time-series are derived. The "standard" model is not robust to data specification. Of the four different specifications tested above, only the heterogeneous error-correction panel allowing for co-integration—i.e., for long-run convergence together with short-run non-homogeneous parameters—partially supports the model, whereas all the homogeneous panel specifications are not supported by the data.

This leads to two considerations. In the first place, the evidence presented in previous studies is statistically flawed. Given the time-series properties of the variables, a simple ordinary least squares (OLS) model estimation would be clearly incorrect. The aggregated one-equation specification of model (1)— all migration to all countries at once—would also be incorrect, as it would not account for the fundamental substitution effect among destination countries: if the salary of country A relative to country B falls, it would be more convenient migrate to country B as opposed to country A. A one-equation aggregated model would also be prone to spurious correlations.[34] So, the various examples of Hatton and Williamson's evidence applied to Italy are to be rejected on a pure statistical ground.

[32] A (small) country might have very high employment rates but a very low demand for labor. It may have very high employment-rate growth rates but a very small size of the economy (think of a country at the beginning of the development process).

[33] Among the various alternative specification we tried, we tested the model with two activity variables constructed by interpolating employment levels with the cumulative growth of GDP. Results are shown in Table S6, the Supplementary material. As it turns out, the estimated model is even less supported than in the previous case, some of the signs are wrong and the variables are not significant.

[34] Consider the simple case of two receiving countries only, A and B, with only one explanatory variable, the wage rate. While the overall (aggregate) migration flow might be growing, it can well be the case that the wage rates might not be

Table 3. *Heterogeneous panel approach – Testing different hypotheses*

| | Malthusian | Poverty trap I | Poverty trap II | Big surge |
|---|---|---|---|---|
| Convergence coefficient | | | | |
| $M_{t-1}$ | −0.035* (−1.82) | −0.124* (−1.86) | −0.042* (−1.06) | −0.101* (−1.69) |
| Long-run coefficients | | | | |
| $\text{Ln}(w_f/w_h)_{t-1}$ | 8.467** (2.16) | −0.013 (0.23) | 0.735 (0.56) | −36.228 (−1.15) |
| $\text{Ln}(e_f)_{t-1}$ | 18.126* (1.80) | 0.101 (1.07) | 1.895 (1.44) | −32.006 (−1.05) |
| $\text{Ln}(e_h)_{t-1}$ | 20.131* (1.68) | 1.635*** (2.85) | −1.554 (−0.28) | −87.370 (1.12) |
| $\text{Network}_{t-2}$ | 0.006** (2.01) | 0.007*** (6.44) | 0.029*** (3.29) | 0.005 (−0.54) |
| $C_{t-1}$ | −0.209 (−0.26) | −0.001 (0.05) | −2.931 (−1.57) | −6.419 (−1.08) |
| Pop Growth | −0.131 (−1.32) | | | |
| GDPpc | | −0.220*** (−2.51) | | |
| GDPpc^2 | | −0.040 (−1.08) | | |
| VA Agr pc | | | 1.559* (1.66) | |
| VA Agr pc^2 | | | 0.126 (0.28) | |
| 1901–1913 | | | | 11.268** (1.89) |
| Number of areas | 11 | 11 | 11 | 11 |
| Observation | 385 | 385 | 385 | 385 |
| Log likelihood | 1041 | 1055 | 1013 | 1011 |

*, **, and ***are 10, 5, and 1 percent significance level, respectively. Dependent variable $\Delta m_t$.

Second, the various panel-model specifications illustrated above model show that, generally speaking, the model is not fully validated, is not robust, and a more satisfactory explanation has to be found for the Italian migration flows observed before the First World War. Moretti (1999) and Faini and Venturini (1994) evidence was also not fully supportive and yet, in their case, one could argue that it is was because of the limited coverage of their migration data—three or four destination countries at the most—and the improper treatment of the data series. This is not our case, as our data cover sixteen countries and we duly account for non-stationarity and co-integration. When we adopt a heterogeneous panel-regression model allowing for both stationarity and non-stationarity in the variables as well as co-integration, results do appear to be more encouraging, albeit still not fully supportive of the model.

We also tried a few tests of specific hypotheses studied in the literature, which have been suggested as possible drivers of migration flows: the Malthusian hypothesis, the poverty-trap hypothesis, the linguistic-homogeneity hypothesis, and the homogeneous-flow hypothesis. These hypotheses can all be tested by alternative specifications of the heterogeneous panel model in equation (5).

The first specification is an augmented version of the basic model above in equation (5) with a population growth-rate variable: this responds to the so-called *Malthusian hypothesis* of a migration flow induced by population pressure (Easterlin 1961b). We tested it by using the "population growth rate recorded 15 years before" as this was the average working age lower limit at the time, as mentioned above. As we see from table 3, however, that particular variable (*Pop Growth*) is not statistically significant and the model is not supported by the data.

increasing, and yet the correlation with the "aggregate" wage rate (a sum) be positive. This can also be shown by looking at the covariance (and the correlations): the covariance of $X$ with $Z + Y$ can be positive even if the covariances of $X$ and $Y$ and $X$ and $Z$ are negative.

The second specification is an augmented version of the basic model with an income variable to account for the existence of a budget constraint. According to this hypothesis, migration is constrained by the possibility to sustain the costs of migrating: as income grows over a certain level, migration becomes possible, up to a level where it would not be necessary, as income is large enough. This, which has been termed as the *poverty-trap hypothesis*, has been studied, among others, by Faini and Venturini (1993, 1996, 2010). In table 3 we show two different specifications, where we used either Italian per-capita GDP—as an indicator of affluence—or Italian agricultural value-added—as an indicator of income generated by the largest sector of employment (table 4).[35]

The poverty-trap hypothesis does not seem to be confirmed by the data, in both specifications. Per-capita GDP and the squared per-capita GDP are both significant, but with a negative sign, both the network variable and domestic employment are significant whereas both the wage differential and the receiving-country's employment level become nonsignificant. We should remember that employment rates are calculated (interpolated) using GDP as a basis and are therefore correlated with per-capita GDP. Wages, too, as calculated by Hatton and Williamson, suffer of the same problem.[36]

The third specification is a test of a possible structural break in 1901. As Gould (1979) pointed out, the change in regulations concerning migration from Italy that occurred in 1901 might have caused a break in the model. We have thus introduced a dummy variable in equation (5) to account for that. As it turns out, the model so specified is completely rejected by the data, as no variable is significant but the dummy itself.

We have also tested a number of more specific hypotheses (table 3) that imply a subdivision of the sample of countries, like that of "linguistic affinity", that of "country vicinity", and, finally, that of "dimensional congruity". The hypothesis of *linguistic affinity* for Italian migrants corresponds to the idea that migrating to a country where the spoken language is of Latin origin has a lower cost. We tested this hypothesis by dividing the receiving countries in two groups and then testing the basic model in equation (5) for the two groups separately. In the "Latin" group, we included Argentina, Benelux, Brazil, Canada, France, and Iberia.[37] The hypothesis does not receive any support in the data and is fully rejected. As for the second hypothesis, we divided the countries in two groups—Europe and the rest of the world—and tested the model on the two groups separately. This hypothesis would also account for some influence of temporary versus permanent migration. This specification seems to work for Europe, but not for the rest of the world, although wage rates are not significant, as is domestic employment. As for the third hypothesis, we divided the countries in two groups—the receiving countries with large migration flows and the others—and tested the model on the two groups separately. This specification seems to work somehow, as all estimated coefficients are similar in magnitude to the overall model but less significant.

In sum, Italian migration seems to be explained by wage differentials, foreign activity levels, network effects—the stock of previous migrants, and not by the level of domestic activity. To a degree, Italian migration remains mystifying. The question is whether the model is not fully supported by the data because of improper model specification or because of poor data quality. As we have seen above, the model is quite sensitive to alternative specifications, to the addition of more

[35] Table 4 reports the tests for the two "poverty-trap" specifications. In the first column, we have the log of per-capita GDP (GDPpc) and the log of per-capita GDP to the square (GDPpc), while in the second column we have the log of agricultural value added (VA) and the log of agricultural value added to the square (VA).

[36] As for the agricultural value added variables, they do not appear to be significant.

[37] So we have two panel regressions: one with six equations and the other with five equations.

Table 4. *Heterogeneous panel approach–Italian migration by groups*

| | Latin vs. others | | Europe vs. Americas | | Small vs. large | |
|---|---|---|---|---|---|---|
| Convergence coefficient | | | | | | |
| $M_{t-1}$ | −0.012** (−2.10) | −0.191* (−1.63) | −0.080 | −0.026*** (−3.17) | −0.191* (−1.86) | −0.102 (−1.04) |
| Long-run coefficients | | | | | | |
| $\text{Ln}(w_f/w_h)_{t-1}$ | 2.603 (0.75) | 0.107* (1.91) | 2.019 (1.44) | 3.097 (1.22) | 0.038* (1.66) | 2.500 (1.46) |
| $\text{Ln}(e_f)_{t-1}$ | −1.333 (−0.62) | −0.022 (−0.47) | 17.543*** (3.58) | 0.103 (0.05) | 0.045** (1.98) | 18.539*** (3.36) |
| $\text{Ln}(e_h)_{t-1}$ | −0.629 (−0.09) | 0.624* (1.71) | 5.259 (0.88) | 18.622 (1.06) | 0.091 (0.83) | 5.071 (0.75) |
| $\text{Network}_{t-2}$ | 0.007 (0.48) | 0.006*** (4.54) | 0.033*** (3.90) | 0.004 (0.52) | 0.007*** (6.68) | 0.031*** (3.38) |
| $C_{t-1}$ | 0.169 (0.16) | 0.003 (−0.38) | −1.488 (−0.76) | −0.019 (−0.43) | −0.000 (−0.01) | −2.160 (−0.95) |
| Number of areas | 6 | 5 | 6 | 5 | 6 | 5 |
| Observation | 225 | 187 | 225 | 187 | 225 | 187 |
| Log likelihood | 661.30 | 427.09 | 790.54 | 303.57 | 1034.41 | 76.06 |

*, **, and ***are 10, 5, and 1 percent significance level, respectively. Dependent variable $\Delta m_t$.

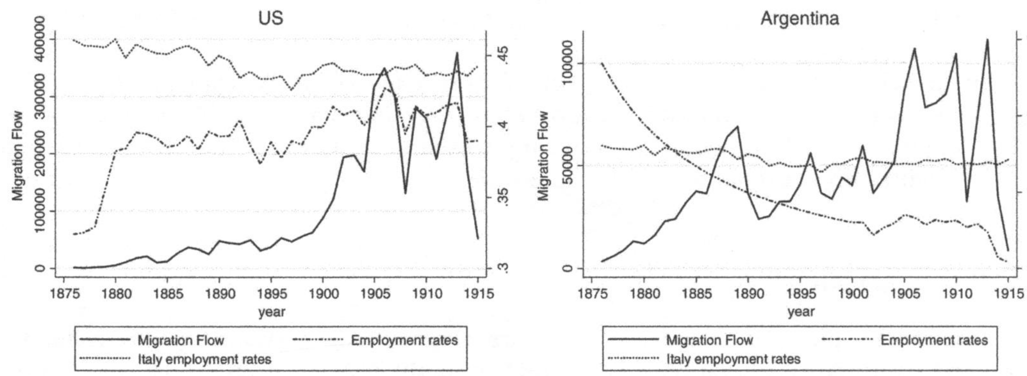


Figure 2. *Italian migration to United States and Argentina.*

variables, and to the sample of countries included. Data quality is always an issue, as is the appropriate data to use for the variables in the model. In the case of wages, for instance, it is true that large shares of the rural labor force were actually not properly *employed* and wage labor was only part of total "employment". Industry in Italy was still quite underdeveloped, and the industrial (construction) wage rate for low-skill jobs might have not been the real alternative for many rural laborers in excess supply. The alternative, for many, was probably between subsistence and anything outside Italy, albeit of an unskilled type. In short, one problem might lie in the wage rate variable we have used for Italy, which comes from Williamson (1995). Either a rural wage or some different composite measures of industrial wages might be more appropriate. A second issue concerns *aggregate* migration data. By considering migration flows to the various countries simultaneously, we put together those who migrated to the United States to work in the mines with those who migrated to Argentina to work in the fields and with those who went to Great Britain to work in the textile industries. Some jobs were permanent, some jobs were seasonal. Some transfers were paid by Governments, others were quite dangerous and risky. Aggregate analyses like this one always average out such differences, and we try to disentangle the long-term fundamental determinants.

That Italian migration outflows to specific countries are quite uncorrelated to domestic employment levels is apparent even at the first sight. Just as an example, consider the migration flows to the United States and to Argentina and how they compare to the Italian employment level (Figure 2). Even when we account for Italian resident population—and we look at migration *rates*, there appears to be no correlation between migration and employment.[38]

Even considering that these are aggregated measures of migration flows, whereby skilled workers are pooled with unskilled ones, it is evident that the (lack of) correlation might therefore be very spurious. In this sense, it would be useful to disaggregate migration flows by profession or sector. Aggregate national data might be a problem, too, as migration flows differed by area and region of origin over time. This point was also stressed by Hatton and Williamson (1998). If domestic employment levels can be taken as an indicator of labor demand—the "push" factor—then there should be a perfectly negative correlation between the flow of emigrants and the employment level at home.[39] The same considerations may apply in this case,

[38] Migration rates as measured in terms of resident population may be taken as indirect measures of excess labor supply.
[39] The same holds for domestic employment growth rates, which are certainly not more correlated with migration flows than levels.

too, whereby more disaggregated data might provide a different picture more in light with our theoretical expectations.

Given this evidence, we may therefore conclude that the simple theoretical explanation—that migration simultaneously responds to wage differentials, higher employment level abroad, and lower employment levels at home—cannot simply hold for Italy as a whole, at the aggregate level, even with a proper treatment of the time-series at hand.

## 5. Conclusions

The main conclusions of the paper can be summarized as follows. What we have called the "standard" quantitative migration model—best exemplified by Hatton (1995) and Hatton and Williamson (1998)— is only partially validated for the Italian migration over the 1876–1913 period. We have estimated eleven migration equations for eleven different destinations simultaneously, using a variety of panel-regression models and yet, the model specification only appears to be partially supported in the heterogeneous case, when accounting for different short- and long-run dynamics across countries of destination. Thus, as the problem does not appear to be one of improper treatment of the time-series properties of the data, we may conclude that there are other reasons for such a result: that we have used and applied the "wrong" data series or that the model specification is inappropriate. As the heterogeneous panel model confirms, it is the heterogeneity in the migration flows—by country, and thus possibly by area or region, by sector of employment, or by social and professional status and what else—that counts. Such heterogeneity has two sides: one that is due to the country of destination and the other that is due to the conditions (region, social and economic status, etc.) at the origin. Aggregate data at the national level are bound to mask such heterogeneity and a further level of disaggregation has to be tried: at the geographical level, migration determinants may be different by area, as emigration greatly varied across regions; at the time level, flows changed over time; and at the level of composition of the migrant pool, not all migrants were equal by economic and demographic characteristics. This is all matter for future research.

Italian emigration must have had specificities that are not accounted for in the standard "aggregate" model. "Italians migrated to a much wider variety of destinations than did emigrants from any other European country", as Hatton and Williamson (1998) stated, which adds to the diversity of migration flows. Also, the answer to Hatton and Williamson (1998) question cited above: "To what extent can Italian emigration be explained by the same economic and demographic forces?" seems to be, in our opinion, "only to a partial extent, there must be something else". Also, "rising European emigration was driven chiefly by natural population increase, industrialization and the rising emigrant stock itself".[40] If there was a growing population, the population surplus, which started much earlier particularly in rural areas, was faced with an agrarian crisis, low productivity in agriculture, low agricultural salaries, and higher food prices, all factors that are not reflected in the aggregate nation-wide variables we have used in testing the model. That "rising per-capita income releases the poverty constraint", as Faini and Venturini (1994) claimed, cannot be taken as a factor for increasing migration. As we have seen above, both the "Malthusian" and the "poverty-trap" hypotheses add nothing to the explanatory power of the model.[41]

[40] Hatton and Williamson (1998, p. 96).

[41] In addition, if the model could not fully account for the large Italian migration flows, the key to the answer might be that it is not a true representation of the increased flow of migrants *settling permanently abroad*, as they also tended to go

This paper is part of a broader research project on the determinants of Italian migration before the Great War and their relationship with Italian economic development. In the paper, we have put to test the standard quantitative economic model of Hatton and Williamson and others, by properly treating the data at hand, using the correct econometric approach and accounting for the heterogeneity of destinations that was so important in determining population movements in the period. What our results show is that by properly accounting for the time-series properties of the data and by using the proper model estimation procedure—an heterogeneous co-integrated panel-regression model—the standard quantitative economic model does have some explanatory power and yet is not fully supported: part of the explanation is still left out. In that specification, we are able to account for the *heterogeneity of destinations*, which is certainly responsible for the different responses to migration flows to different countries over time. However, that specification does not account for the *heterogeneity of origins*, which is possibly responsible for the "left out" part of the explanation. Some of the migrants came from poor internal rural areas in Italy, some others came from the cities. Some migrants were landowners looking for a job that would give them an additional income to accumulate and go back home, some others were wage laborers who did not plan do go back. In short, by aggregating over migration flows by origin (in Italy) and by profession and age and other demographic characteristics, we lose the nuances and flatten out the differences. The heterogeneity that appears to be so important for the receiving countries is thus possibly going to be an important explanatory factor of the differences in migration flows from the different parts of Italy to different countries at different times.

We will have to look at the data at a more disaggregated level, accounting for the segmentation across Italian regions and across different destinations, i.e., the existence of segmented labor markets and segmented emigration streams. This will bring to the fore the different motives and profiles of Italian emigrants, and the connections with temporary emigration and high return rates. What this re-visitation of the standard model has shown, in our opinion, is that to explain the Great Italian Migration, there is a great deal of heterogeneity that is not fully taken into account by only looking at the destinations side. Both the heterogeneity of destinations and the heterogeneity of origins have to be accounted for simultaneously. These are all directions that future research will have to explore to come up with a convincing, robust, and reasonable explanation of Italian emigration in the period before the Great War. A first step was needed, yet: testing the standard aggregate model and see how much of the overall flows it can explain. Not much, not all, it seems.

**Acknowledgements**

Earlier versions of the paper have been presented in a number of seminars and benefited from the comments by various colleagues, most notably Maria Elena Bontempi, Matteo Cervellati, Roberto Golinelli, and Vera Negri Zamagni, together with three anonymous referees, to whom the authors are deeply indebted. The research was made possible thanks to the funding from the Department of Economics on "The Determinants of Early Italian Development".

back. As a matter of fact, an estimated third of all Italian emigrants went back home, i.e., they did not settle abroad. "Net migration", i.e. the difference between the outflow and the inflow, fell after 1900. "Italian emigration was accompanied by a rising tide of return migration", pointed Hatton and Williamson (1998): of course, Italians emigrated in order to earn money and go back, not in order to settle. So this was already a difference in motivations.

## Supplementary material

Supplementary material is available at *EREH* online.

## References

BLACKBURNE, E.F. and FRANK, M.W. (2007). Estimation of nonstationary heterogeneous panels. *Stata Journal* 7, pp. 197–208.

Commissariato Generale dell'Emigrazione. (1925). *Annuario Statistico della Emigrazione*. Roma: Serie.

DEL BOCA, D. and VENTURINI, A. (2005). Italian migration. In K.F. ZIMMERMANN (ed), European Migration: What Do We Know? Oxford: Oxford University Press.

EASTERLIN, R. (1961a). Regional income trends, 1840–1950. In S. HARRIS (eds), *American Economic History*. New York: McGraw-Hill, pp. 525–47.

EASTERLIN, R. (1961b). Influences in European overseas emigration before World War I. *Economic Development and Cultural Change* **9**, pp. 531–50.

FAINI, R. and VENTURINI, A. (1993). Trade, aid and migrations: some basic policy issues. *European Economic Review* **37**, pp. 435–42.

FAINI, R. and VENTURINI, A. (1994). Italian emigration in the pre-war period. In T. J. HATTON and J. G. WILLIAMSON (eds), *Migration and the International Labor Market 1850–1939*. London: Routledge, pp. 85–94.

FAINI, R. and VENTURINI, A. (1996). Migration and growth: the experience of Southern Europe. *CEPR Discussion Paper* 964. London: CEPR– Centre for Economic Policy Research.

FAINI, R. and VENTURINI, A. (2010). Chapter 5: Development and Migration: Lessons from Southern Europe. In G.S. EPSTEIN and I.N. GANG (eds), *Migration and Culture* (Frontiers of Economics and Globalization, Volume 8). Bingley: Emerald Group Publishing Limited, pp.105–136.

FERENCZI, I. and WILLCOX, W.F. (1929). *International Migrations. Vol. I Statistics*. New York: National Bureau of Economic Research.

FOERSTER, R.F. (1919). *The Italian Emigration of Our Times*. Cambridge, MA: Harvard University Press.

GOMELLINI, M. and O'GRADA, C. (2011). Outward and Inward Migrations in Italy: A Historical Perspective. *Bank of Italy, Quaderni di storia economica* (Economic History Working Papers) 8.

GOULD, J.D. (1979). European intercontinental emigration, 1815–1914: patterns and causes. *Journal of European Economic History* **8**, pp. 593–679.

HARRIS, J.R. and TODARO, M.P. (1970). A two-sector model of migration with urban unemployment in developing countries. *American Economic Review* **60**, pp. 126–42.

HATTON, T.J. (1995). A model of UK emigration, 1870–1913. *Review of Economics & Statistics* **77**, pp. 407–15.

HATTON, T.J. and WILLIAMSON, J.G. (1993). After the famine: Emigration from Ireland. *Journal of Economic History* **53**, pp. 575–600.

HATTON, T.J. and WILLIAMSON, J.G. (1994). *Migration and the International Labor Market 1850–1939*. London: Routledge.

HATTON, T.J. and WILLIAMSON, J.G. (1998). *The Age of Mass Migration*. New York: Oxford University Press.

MACDONALD, J.S. (1958). Some socio-economic emigration differentials in rural Italy, 1902–1913. *Economic Development and Cultural Change* **7**, pp. 55–72.

MACDONALD, J.S. (1963). Migration and labour militancy in rural Italy. *Economic History Review* **16**, pp. 61–75.

MACDONALD, J.S. and MACDONALD, L.D. (1964a). Chain migration ethnic neighborhood formation and social networks. *Milbank Memorial Fund Quarterly* **42**, pp. 82–97.

MACDONALD, J.S. and MACDONALD, L.D. (1964b). Italian migration to Australia: manifest functions of bureaucracy versus latent functions of informal networks. *Journal of Social History* **3**, pp. 249–75.

MITCHELL, B.R. (1993). *International Historical Statistics: The Americas, 1750-1988*. New York: Macmillan.

Moretti, E. (1999). Social networks and migrations: Italy 1876–1913. *International Migration Review* **33**, pp. 640–57.
Pesaran, M.H., Shin, Y. and Smith, P. (1999). Pooled mean group estimation of dynamic heterogeneous panels. *Journal of Econometrics* **68**, pp. 79–113.
Ravenstein, E.G. (1889). The laws of migration. *Journal of the Royal Statistical Society* **52**, pp. 241–305.
Rosoli, G. (1978). *Un Secolo di Emigrazione Italiana 1876–1976*. Roma: Centro Studi Emigrazione.
Sori, E. (1979). *L'Emigrazione Italiana Dall'Unità Alla Seconda Guerra Mondiale.* Bologna: Il Mulino.
Vecchi, G. (2011). *In Ricchezza e Povertà*. Bologna: Il Mulino.
Williamson, J.G. (1995). The evolution of global labor markets since 1830: Background evidence and hypotheses. *Explorations in Economic History* **32**, pp. 141–96.