The missing ingredient: distance internal migration and its long-term economic impact in the United States

The missing ingredient: Distance

Internal migration and its long-term economic impact in the United States

Viola von Berlepsch and **Andrés Rodríguez-Pose**

Department of Geography and Environment
London School of Economics
Houghton St
London WC2A 2AE, UK
e-mails: v.k.von-berlepsch@lse.ac.uk; a.rodriguez-pose@lse.ac.uk

Abstract:

This paper examines if internal migrants at the turn of the 20th century have influenced the long-term economic development of the counties where they settled over 100 years ago. Using Census microdata from 1880 and 1910, the distance travelled by American-born migrants between birthplace and county of residence is examined to assess its relevance for the economic development of US counties today. The settlement patterns of domestic migrants across the 48 continental states are then linked to current county-level development. Factors influencing both migration at the time and the level of development of the county today are controlled for. The results of the analysis underline the economic importance of internal migration. Counties that attracted American-born migrants more than 100 years ago are significantly richer today. Moreover, distance is crucial for the impact of internal migration on long-term economic development; the larger the distance travelled by domestic migrants, the greater the long-term economic impact on the receiving territories.

Keywords: Internal migration, distance, long-term, economic development, counties, US

JEL Codes: J61, N11, O15, R23

1. Introduction

"For we began as explorers, empire builders, pilgrims and refugees, and we have been moving, moving ever since." George Wilson Pierson (2011:91)

In 2015, the United Nations estimated that there were 244 million international migrants worldwide (United Nations, 2016). The size of global internal migration, however, was more than three times larger: around 763 million people or 11.7 percent of the world's population (Kuhn, 2015). Yet, despite this imbalance in size, our knowledge about the economic impact of domestic migration is more limited than that of international migrants. Measurement difficulties coupled with the relative lack of popularity of research on internal movements, have kept the study of the economic impact of domestic migrants pinned to the bottom of the academic agenda.

Past research has focused mainly on the determinants and patterns of internal migration, the migrants' characteristics, the individual returns of the migration decision and of remittances, but the macroeconomic impact of internal migrants on their receiving region remains poorly understood. The few insights on this topic focus on the short- to medium-time frame, barely extending further than 10 to 20 years. The understanding of the long-lasting economic impact of domestic migration is virtually non-existent. Is the settlement pattern of historical internal migrants relevant for current disparities in economic development? Has domestic migration taking place more than 100 years ago left an influence that can still be traced today?

The paper intends to cover these gaps in our knowledge by focusing on a country often considered as an exceptional case of internal migration – the United States (US). US citizens have traditionally displayed geographical mobility rates nearly double those of other advanced societies; almost one-third of American-born citizens lived outside their state of birth (Molloy et al., 2011; US Census Bureau, 2016). The annual geographical mobility rate in the US has historically been about three times larger than that of the EU15, EU27, or Canada and about one and a half times that of Australia (Gill and Raiser, 2012). This high internal migration is a legacy of the country's history. Already throughout the 19th century almost 60% of the US-born male population over the age of 30 had moved across county- or state-lines (Ferrie, 2005). By 1880, more than a third of the US population were American-born internal migrants (own calculations).

Using a county-level dataset for domestic migration[1] in the late 19th and early 20th century covering the 48 US continental states, the paper first assesses the effect of historical internal migration on long-term economic development. It evaluates whether large shares of domestic migrants have left a long-lasting trace on the territories where they settled. Second, it examines whether the distance covered by American-born migrants more than a century ago matters for current levels of development.

We assume that a vibrant domestic migrant population can galvanise growth over the long-term, leaving a long-lasting economic impact, determined by the geographical distance covered by internal migrants. Large numbers of internal migrants travelling over large distances can transform the counties of destination, by increasing diversity levels, transforming local institutions, and reshaping economic activity. The hypothesis driving the paper is that the distance covered by internal migrants affects future local economic performance: the bigger the distance travelled by internal migrants more than 100 years ago, the larger the differences between new arrivals and locals, the higher the population diversity, the greater the local economic dynamism, and the stronger the long-term economic legacy.

To demonstrate whether this is the case, the paper adopts the following structure. First, the historical background of internal migration in the US at the turn of the 20th century is described in section 2. Section 3 summarizes the literature on internal migration and economic development, while section 4 provides a description of the empirical approach and the data used in the analysis. The discussion of the results can be found in section 5. Section 6 concludes.

2. Internal migration at the turn of the 20th century

Since its foundation, the US has been characterised by an exceptionally high population mobility. Rapid economic expansion in the late 18th and 19th centuries was fundamentally the result of increased supply in two production factors: land and labour. Over the span of a few decades, the US territory expanded 3.5 times to encompass around 7.8 million km^2 by 1900. Population grew almost forty-fold over the same timeframe, as a consequence of both natural growth and immigration (Gallman, 2000).

[1] In this paper the terms *domestic migrants* and *internal migrants* refer to American-born population moving from one place to another and are used as synonyms. The terms *American-born* and *native-born* both refer to population born on American soil, rather than to the native indigenous populations.

During this period, geographic mobility increased to previously unforeseen levels. Ferrie (2005) estimates that nearly two thirds of American-born men above the age of 30 migrated across county- or state-lines during their lifetime. By 1850, the share of American-born population living outside their place of birth was almost 25 percent (Haines, 2000). By 1880, it reached 33 percent (Ruggles et al., 2015; own calculations). "The American population was a restless one, continually uprooting and moving to a new location […] 'every day was moving day'" (Atack and Passell, 1994: 237).

Table 1 portrays the population redistribution over time. In 1790, no population is recorded in the Midwest.[2] By 1830, already 12.5 percent of the US population had settled in this area and by 1860 one third of the entire population lived in the region. The West Coast grew at a similar rate, only later. In 1860 half a million people lived in the Pacific and Mountain states. By 1890 20 percent of the American population lived there and, by 1910 it was almost 30 percent. The geographical centre of the US population was continuously shifting westward (see e.g. Plane and Rogerson, 2015).

Table 1. Population by region, 1790-1910 (in thousands)

	1790		**1830**		**1860**		**1890**		**1910**	
Region	N	%	N	%	N	%	N	%	N	%
New England	1.009	25,7	1.955	15,2	3.135	10,0	4.701	7,5	6.553	7,1
Middle Atlantic	959	24,4	3.588	27,9	7.459	23,7	12.706	20,2	19.316	20,9
East North Central	–	–	1.470	11,4	6.927	22,0	13.478	21,4	18.251	19,8
West North Central	–	–	140	1,1	2.170	6,9	8.932	14,2	11638	12,6
South Atlantic	1.852	47,1	3.646	28,3	5.365	17,1	8.858	14,1	12.195	13,2
East South Central	109	2,8	1.816	14,1	4.021	12,8	6.429	10,2	8.410	9,1
West South Central	–	–	246	1,9	1.748	5,6	4.741	7,5	8785	9,5
Mountain	–	–	–	–	175	0,6	1.214	1,9	2634	2,9
Pacific	–	–	–	–	444	1,4	1.920	3,0	4449	4,8
Total	**3.929**	**100**	**12.861**	**100**	**31.444**	**100**	**62.979**	**100**	**92.228**	**100**

Data source: US Bureau of the Census, 1972

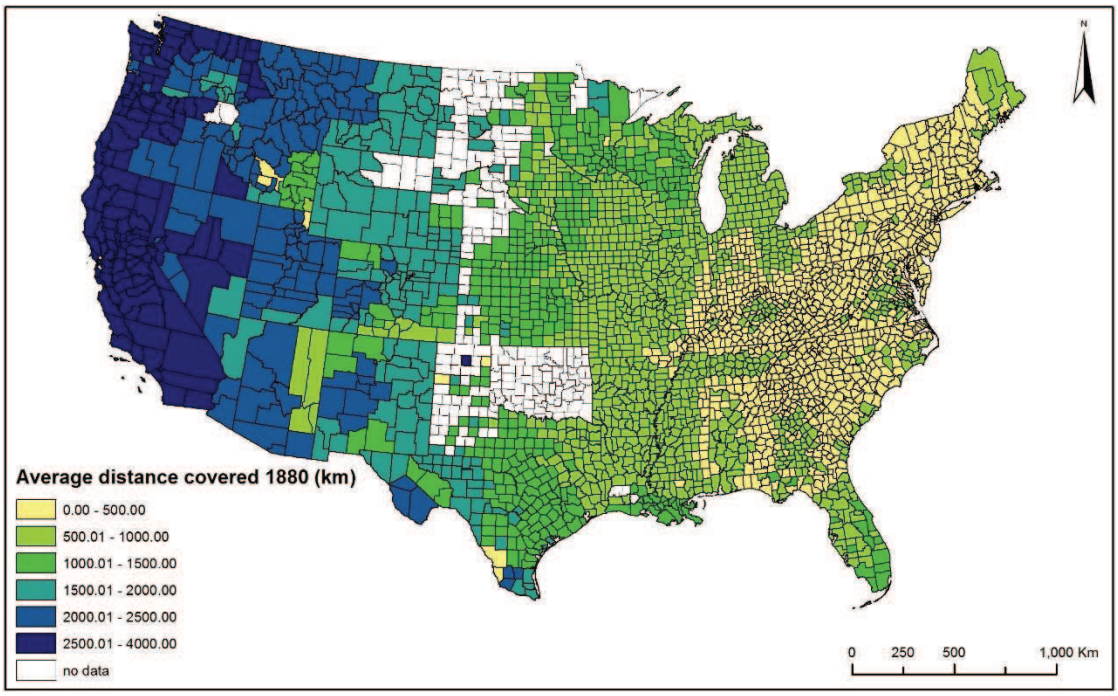
Settlement patterns were far from random. Topography, climate and natural amenities determined settler flows (Merck, 1978). Appendix 1 displays the settlement pattern of domestic migrants, measured as the share of a county's total American-born resident

[2] The history of American geographical mobility ultimately starts with native Americans. Historians estimate more than around 143 'language families' lived on the territory to the north of Mexico prior to the arrival of the Europeans (Spickard, 2007). Thus, the American continent was far from 'empty' and American expansion to the West was by no means a peaceful undertaking. Unfortunately, as data on the settlement patterns or economic fundamentals of the population living in these areas and, especially for the native-American tribes, were not adequately recorded, we are unable to include this important part of the American history into the analysis.

population in 1880. A clear east-west divide emerges. The lowest numbers of internal migrants are found in the original thirteen states. The share of domestic migrants rises rapidly toward the West. In 1880, states in the Midwest and westerm mountain regions had the largest proportion of internal migrants.

In the late 19th century most domestic migrants travelled short distances (i.e. Ravenstein, 1885, 1889): in 1880, 80 percent of all internal migrants settled less than 500 km away from their place of birth. For example in 1850, 67 percent of Pennsylvanian-born migrants lived in Ohio, Indiana, or Illinois; 77 percent of South Carolina migrants settled down in Georgia, Alabama, Mississippi, and Tennessee (Haines, 2000). Ravenstein (1885, 1889), one of the first to analyse migratory patterns, explained these short-distance movements by differentials in economic development. People were drawn by factors in close-by regions that allowed to improve their economic prospects. The remaining 20 percent covered much larger distances, in some cases up to 4,500 km (Ruggles et al., 2015; own calculations). Figure 1 displays the average distance travelled by domestic migrants in 1880. The further west, the larger the distance covered. Distances were shortest in the Northeast, exceeded 500 km in the Midwest, and reached 1,500 km in the western part of the Great Plains. The distance travelled by an internal migrant living on a Pacific coast county ranged between 2,500 and 4,000 km.

Figure 1. Average distance travelled by American-born migrants 1880 (in km)



Data source: Ruggles, et al. 2015; own calculations

The migrant catchment areas varied widely in size: in 1880 the average internal migrant living in New York City was born an average of 460km away; an internal migrant living in Harris County (Houston, Texas) had travelled 1,200 km; while in San Francisco City s/he had covered more than 3 times the distance – 3,700 km (Appendix 2).

3. Internal migration and economic development

Uncovering the link between internal migration and economic development is difficult. Researchers analysing domestic migration face two challenges: the inconsistent and difficult conceptualisation and subsequent measurement of internal migration, on the one hand, and its limited popularity within economic research, on the other. Problems related to defining, measuring, and collecting data on domestic flows have hindered research on internal migration. Many different conceptualisations of internal migration, employing a variety of temporal intervals and spatial entities, have been coined (i.e. Petersen, 1986; Rees et al., 2000; Rogers et al., 2003). However, "the size and shape of the spatial areas between which migration is measured are not uniform either within or between countries" (Greenwood, 2015: 443). Nor is there an internationally agreed standard for measuring internal migration. Drawing robust conclusions, identifying implications, or even calculating internal migration flows is hence a difficult undertaking. The research field to estimate the economic impact of internal migration is, as a consequence, limited in size (i.e. Kupiszewska and Nowok, 2008; Bell et al., 2015) and there are few exceptions to this norm. Most of the exceptions have come from historians, such as Gregory (2005) or Wilkerson (2010). Gregory (2005) stresses that internal migration has been underestimated as a driver of economic change in American history. Tolnay et al. (2005) addressed the question of the role of distance for racial selectivity and spatial sorting of black and white internal migrants, as well as that of European immigrants in Chicago. However, on the whole, the popularity of migration flows across international borders has detracted attention from internal migration. Even though the bulk of global and US geographical mobility takes place within national borders, the analysis of international migration has dominated economic research and policy discussions (Skeldon, 2006; Ellis, 2012). Studies on "population movements involving changes of residence within countries remain poorly developed" (Bell, et al., 2015:33).

Often, the word 'migration' has been altered in meaning to refer almost exclusively to international migration. Internal migrants are frequently subsumed under a 'residential mobility' or 'population distribution' category (Skeldon, 2006). The interest in internal

migration is simply not proportional to the amount of people it concerns and remains "relatively understudied, undermeasured, and misunderstood" (Kuhn, 2015: 433).

Much of the substantive literature delving into the economic impact of internal migration dates from the 1980s and earlier (i.e. Ravenstein, 1885, 1889, Thomas, 1936, 1941; Gordon, 1954; Thomas and Kuznets, 1957; Lee, 1966; Åkerman, 1975). This literature focuses predominantly on the patterns, streams, and changes across time in internal migration. Emphasising in particular the move between agricultural and urban as well as industrial areas, this research concentrates on social change as well as on path dependency. Migrants moved to their places of destination following opportunity (Ravenstein, 1885, 1889; Lee, 1966) meaning that internal migration often bolstered the pre-existing advantages of the places of destination in a Myrdalian 'cumulative causation' process. The places of origin, by contrast, lost out, contributing to a greater territorial polarization of the country.

The insights gained from research on internal migrants can be grouped into three research strands – the individual level, the regional perspective, and migratory patterns. Research on the individual migrant has concentrated on the determinants of migration, the migrants' characteristics and individual returns. Individual features, such as age, schooling, marital status, health, job tenure, poverty or employment status, earnings, or retirement status have been analysed mainly in the context of individual utility maximization models (Plane, 1993; Greenwood, 1997; Bell et al., 2015). Generally, internal migration is found to lead to positive individual returns in earnings and employment (Yap, 1976; Molloy et al., 2011), without necessarily improving subjective well-being (De Jong et al., 2002; Sloan and Morrison, 2016).

A second group of studies adopts a regional perspective. Many focus on movements between rural and urban areas, evaluating social costs, brain drain, and integration issues (Huning and Huetl, 2012; Rupasingha et al., 2015). Place-specific pull- and push-factors, such as the employment rate, wage differentials, tax rates, public services, local government quality, social capital, climate, and other local amenities are also covered (i.e. Delisle and Shearmur, 2010; Biagi et al., 2011; Kuhn, 2015).

The final group has evaluated the size, patterns, and trends of internal geographical mobility. The initial work of Ravenstein (1885, 1889) and Lee (1966) is at the origin of subsequent research by Borjas et al. (1992), Newbold and Bell (2001), or Molloy et al. (2011). Internal migration rates have been calculated and compared across time and space, leading to policy recommendations (Aking and Dökmeci, 2015).

The macroeconomic consequences of internal migration and their economic impact, however, remain – to the best of our knowledge – overlooked by recent scientific research (White and Lindstrom, 2005). Contemporary studies on domestic migration, moreover, deliver inconsistent results. For some, internal migration leads to higher incomes, lower inequality, modernization, and growth (e.g. Yap, 1976; Berker, 2011; Kuhn, 2015). Other studies show that domestic population movements result in regional divergence and widening inequality, significantly lowering growth (Molloy et al., 2011). A third group fails to find any robust relationship between both factors (White and Lindstrom, 2005). In short, the limited research and wide range of findings in the literature on the macroeconomic impact of domestic migrants leaves us with no clear answers as to how internal migration shapes the economic growth trajectory of the receiving regions.

In order to establish a hypothesis about the potential link between internal migration and economic development, we therefore need to resort to the literature on international migration (Ellis, 2012). Pryor (1981), King and Skeldon (2010), and Ellis (2012) question the dichotomy in the field of migration studies between internal and international migration pushing towards an integrated system embracing the similarities: when analysing internal population flows in the simplest way, they could be defined – from a basic labour economics standpoint – as "a major mechanism through which labour resources are redistributed geographically" (Greenwood, 1997:648). If we assume that the economic effect of internal migration mirrors that of international migration, domestic migrants will have a generally economic growth enhancing effect (Borjas, 1995; Card, 2005). The transmission channels identified by the international migration literature include the mere expansion of the labour force (Ortega and Peri, 2009), increasing returns to scale (Borjas, 1995), adjustments in the local market's skill- and labour-composition (Lundborg and Segerström, 2002), increases in wages (Ottaviano and Peri, 2006), and the stimulation of productivity by means of innovation and skill-set extensions (Alesina and La Ferrara, 2005; Gordon and McCann, 2005; Partridge and Furtan, 2008). Further growth potential can be derived from the increased population diversity as a result of labour inflows (Jacobs, 1969; Florida, 2002; Saxenian, 2006; Rodríguez-Pose and von Berlepsch, 2018). Consequently, internal migrants are expected – similarly to their international counterparts – to positively affect the growth trajectory of receiving territories.

Traditional migration literature emphasises that one of the differentiating factors between international and internal migrants is the level of diversity in the receiving region. As

migration research defines diversity mostly referring to birth countries, ethnicities, or languages spoken (i.e. Alesina and La Ferrara, 2005), internal migrants are generally considered not to increase population diversity. Domestic migrants are native-born. They share the same birth country – and often ethnicity and language – with the local population in the receiving region. But these similarities do not necessarily mean that internal migrants do not enhance diversity. We beg to differ in this regard. Greater diversity within societies welcoming internal migrant inflows is related to the distance covered by internal migrants. Internal migrants come from the same country, but often from faraway cities and regions, frequently with different habits and customs. While an Oregonian in Washington State will have had a short trip, a New Yorker in the same place would have covered more than 2,500 miles. This geographical distance and the diversity in places of origin – in spite of the fact that both domestic migrants speak the same language, share the same country of birth and, possibly, ethnic traits – distinguishes them. Hence the distance travelled represents an indicator of (dis)similarity in the institutional baggage internal migrants bring – individual traditions, customs, habits, and different mind-sets are shaped by the place of birth. In his path breaking work on Italy, Putnam (1993) demonstrates that institutional constructs are highly place-specific and vary greatly from one region to another. The closer two regions are, the greater the similarity between institutional constructs; the further away, the greater the difference (Arbia et al., 2010). A New Yorker in Washington State – especially in historical times – will have brought institutional constructs significantly different from those of the local Washingtonian. The institutional baggage brought by an Oregonian would have been less different. A New Yorker in Washington State would therefore have raised population diversity levels to a greater extent than an Oregonian.

In brief, the geographical distance travelled by the internal migrant will affect population diversity in the areas of destination, shaping their economic prospects. The greater the distance travelled by migrants, the greater the population diversity at destination, and, consequently, the higher the growth prospects of receiving territories. Large numbers of American-born settlers covering long distances would have contributed to transform the economic fortunes of receiving areas and, through their influence on local diversity levels, positively affected growth (Rodríguez-Pose and von Berlepsch, 2018).

Geographical distance *per se* has, however, only rarely been considered in migration research. Only a few studies mention and/or calculate the distance travelled between place of origin and destination (e.g. Tolnay et al., 2005; White and Lindstrom, 2005). Data and definition

inconsistencies coupled with accuracy issues of distance measures hamper the retrieval of robust results (Niedomysl et al., 2017). Geographical distance has mostly been connected to migration when evaluating long-distance in comparison to short-distance migration (Ravenstein, 1885, 1889). Dynamics, characteristics of migrants, racial selectivity, and causes of migration were found to vary widely with distance (Tolnay et al., 2005; Biagi et al., 2011; Pendakur and Young, 2013; Niedomysl and Fransson, 2014) and most studies report that distance is negatively correlated to the size of migration (i.e. Greenwood, 1997; Hipp and Boessen, 2016). Furthermore, while long-distance moves are usually motivated by economic opportunities, short-distance moves are more correlated with improvements in quality of life (Ravenstein, 1885, 1889; Niedomysl, 2011).

In addition, past research has put the emphasis on the short- to medium-term impact of both internal and international migration. The focus has been on the immediate economic effects, covering a maximum of two decades since arrival of the migrant. Whether or how past migration affects regional economic performance after these initial years remains a black box. Apart from predominantly historical studies (e.g. Gregory, 2005; Tolnay et al., 2005; Wilkerson, 2010), Rodríguez-Pose and von Berlepsch (2014) or von Berlepsch et al. (2019) are among the few exceptions seeking to formalize the long-term impact of migration within economic analysis. They find that migration in the 19th century US improved the long-term development prospects of recipient areas.

The persistence of growth enhancing effects of migration over the very long-term is often associated with the role of institutions in recipient areas (Acemoglu et al., 2001; Duranton et al., 2009; Tabellini, 2010). According to these studies, migrants convey their institutional constructs, "[…] in the way of culture, religion, social networks and links with the society of origin" (Joly, 2000:30) from their place of origin to the destination region, preserving customs, traditions, habits, and mentality. As Putnam (1993), Acemoglu et al. (2001), or Duranton et al. (2009) indicate, institutional frameworks persist in time, becoming engraved in the territory. Rodríguez-Pose and von Berlepsch (2014, 2015) theorise that the institutional frameworks derived from 19th century migration inflows into the US still shape current economic growth in the recipient areas. Whether the diversity of institutional constructs brought to places of destinations by short- and long-distance internal migrants has created a similarly positive and long-lasting economic effect remains an open question.

In this article, we tackle the aforementioned gaps in internal migration research by establishing a connection between shares of domestic migrants in a regions' population, the

distance covered by migrants, and economic development over the long-term. Two different research questions are examined: a) Do internal migrants shape long-term economic development in the same way as external ones? b) Does the distance covered by migrants matter for the influence they have on the subsequent growth of receiving regions?

4. Empirical Approach

The Model

We estimate two different models to answer our research questions. Model 1, focusing on different migrant population shares, takes on the following form:

$$y_{i,t} = \alpha + \beta Mig_{i,t_0} + \partial X_{i,t-k} + \theta Z_{i,t_0} + \mu state + \varepsilon_{i,t} \quad (1)$$

Where y is the natural log of income per capita of county i in period t (t=2010); *Mig* is our main variable of interest representing different specifications of the migrant population composition in a given county i at t_0 (t_0=1880 or 1910); X represents a vector of factors linked to income per capita levels of county i at time t-k (k=10), and Z corresponds to a vector of similar factors associated to the level of economic development of county i at time t_0, shaping the attractiveness of the county at the time of migration. Lastly, *state* represents state specific fixed effects, controlling for arbitrary spatial correlations between counties within any given state, and ε describes robust standard errors.

We estimate Model 1 in five different specifications, each using a variation of *Mig*. We first run (1) focusing on the share of total migrants – internal and external (foreign-born) – in a given county i at time t_0 (specification 1) measured as percentage of total county population. Subsequently, (1) is estimated distinguishing between the share of domestic migrants (*IntMig*), measured as the percentage of American-born residents having crossed state-lines between their birthplace and place of residence, and the share of external migrants (*ExtMig),* measured as the percentage of foreign-born in a county's population, as our two variables of interest (specification 2). The reason for controlling for both domestic and foreign-born migrants as their settlement patterns were often shaped by different factors. While domestic migrants would have had fewer constraints on mobility, external migrants mostly arrived in the US through specific gateways – such as Ellis Island in New York – that influenced their subsequent mobility within the country, creating a qualitative sorting factor that parallels the distance-factor for internal migrants. Specifications 3, 4, and 5 focus on different American-

born groups in county *i*'s population at time t_0 while controlling for the share of external migrants. The American-born county population is divided into 'stayers', i.e. those born in the same state (specification 3), internal migrants from any neighbouring state (specification 4), and internal migrants from any other non-adjacent state (specification 5). Each group is measured as the percentage of total American-born population living in the county.

Model 2 estimates the direct relevance of the distance travelled by migrants on economic development 100 to 130 years later. Dependent and control variables remain the same as in Model 1. The independent variable of interest, however, changes to *Distance*, representing the average distance travelled by all migrants – both external and internal – of a given county *i*'s population at time t_0 between their birthplace and county of residence (specification 1) and the distance covered by domestic migrants – American-born residents having crossed state-lines – only (specification 2). The model takes the following form:

$$y_{i,t} = \alpha + \gamma Distance_{i,t_0} + \partial X_{i,t-k} + \theta Z_{i,t_0} + \mu state + \varepsilon_{i,t} \qquad (2)$$

The data

The migration dataset used for the main variables of interest in Model 1 was constructed from the Integrated Public Use Microdata Series USA database (IPUMS) Version 6.0 (Ruggles et al., 2015). This database comprises representative population-weighted samples covering US Censuses and American Community Surveys between 1850 and 2015.[3] Starting from 1850, the US census included information on individual birthplaces, noted as either the country of birth for the foreign-born or the state of birth for the domestic population. Complete birthplace and residence data for all continental US states are available only from 1880 onwards, allowing to trace 'lifetime migration' (put simply, the migration between an individual's birthplace and his place of residence at the time of the census). 1880 is selected as base year for the analysis – a 130-year timeframe between dependent and main independent variable of interest. The population data variable for 1910 is chosen to test the robustness of the 1880 results. As in 1910, the population and industry structure across the US already differed considerably from that of 1880, the 1910 sample represents an ideal robustness test of the 1880 results, while keeping the long-term dimension. Unfortunately, neither inter-county or inter-state moves, nor the date of the location change, nor the duration of residence were recorded prior to 1935. These factors can thus not be incorporated into the analysis.

[3] The American Community Survey was only initiated in 2005.

The sample includes 5,791,531 individuals in 1880 – representing 11.5 percent of the total US population at the time – and 923,153 in 1910 – 1 percent of the population. All individual data were matched to the specific county of residence and aggregated at the county level. All US residents not born in the US are defined as 'external migrants', while all American-born with a birthplace different from their state of residence are classified as 'internal or domestic migrants'.

Due to changes in size, geography, and quantity of US counties over the period of analysis, counties in 1880 and 1910 were matched to their 2010 equivalent using cartographic boundary files provided by the US Census. All county boundaries were normalised to their 2010 borders. Historical county averages were calculated and weighted by the population density at the time of the boundary change. This method allowed us to attribute historical county features to all counties of the 48 continental US states in 2010 (with the exception of 1880 values for Oklahoma).[4]

The second set of independent variables of interest – the average distance covered by the migrant population living in a given county – was constructed using GIS software calculating the point-distance matrix between the centroids of all US counties of residence and the centroids of the 48 continental states (weighted by the population density at the time of migration) as well as all countries named as birthplaces by the foreign-born population. The individual distance travelled by each migrant was allocated to the county of residence and aggregated at county level. Two different specifications – average distance covered by all migrants (both external and internal) and average distance covered by internal migrants only – are calculated. As within-state migration is not recorded in the data, all individuals in a county born in the same state as their county of residence are assumed to either having moved only across county lines or not at all.

The dependent variable (income per capita levels in 2010 in US Dollars) as well as vector X, containing data for 2000, employ information extracted from the US Bureau of Economic Analysis (BEA) database, the Current Population Survey (CPS) tables of the US Bureau of Labor Statistics (BLS), and the 2000 Census Summary files. Vector X is measured 10 years before the dependent variable and is included in the model to control for county-level characteristics directly affecting the level of income per capita of a given US county. X is included as a means of avoiding that recent county features could transfer their individual association to current levels of economic development onto the variable of interest. This

[4] Oklahoma only became an organized territory in 1890.

would potentially lead to over- or underestimating its effect. The factors considered include educational attainment (share of people with college education), the unemployment rate, the share of black and female population, overall population size (as natural log), the share of the labour force employed in manufacturing, the infant mortality rate as a proxy for levels of public health, and the women's participation rate in the labour force. All variables in vector *X* are lagged by 10 years with respect to the period considered in our dependent variable to minimize problems of simultaneity between county features and income per capita. The descriptive statistics for all variables appear in Appendix 3.

Vector *Z*, the second vector of control variables, refers to 1880 and 1910, respectively. It consists of county features that may have influenced the county's level of economic prosperity at the time of migration. These characteristics served as pull-factors to migrants determining the level of attractiveness of a county in that period. The inclusion of *Z* follows a similar reasoning as X. By controlling for county features at the time of migration, we extract their potentially confounding influence on the regression results from the error term and include it explicitly in the regression. Their potential impact on economic development today can therefore not be transferred onto the internal migration variables. Vector *Z* includes, whenever possible, the same variables as vector *X*. Educational attainment, however, is measured as the literacy rate. We also control for the initial county average income at the time of migration. As income per capita data were not collected at the time, a proxy is constructed with individual data on median total income per occupation in 1950 dollars. All of the historical variables are built using the IPUMS USA and the Inter-University Consortium for Political and Social Research databases (ICPSR). Individual data are aggregated at the county-level, employing the same method used when constructing the migration variables. A description of all variables and sources is provided in Appendix 4.

5. Analysis

Origins of migrants

The first part of the analysis focuses on whether internal migrants in the late 19th century left a significant and long-lasting impact on the counties of destination. Model 1 is employed using ordinary least squares (OLS) Table 2 displays the OLS results of Model 1 specifications (1) through (5) for 1880, each time altering the main variable of interest. Two robustness checks to validate our findings are conducted: (a) an instrumental variable (IV) estimation is

undertaken in Appendix 5 and (b) we shift the base migration year by 30 years to 1910 and rerun both the OLS as well as the IV specifications. The latter results are presented in Appendices 6 and 7.

In line with expectations, migration is positively associated with income per capita levels in specification (1). A large share of migrants in 1880, regardless of birthplace, is strongly and positively connected to county-level GDP per capita in 2010. When splitting the migrant stock of a county into internal and external migrants (regression 2), the coefficients for both types of migrants remain positive and highly significant. Hence, counties that attracted large inflows of US and foreign settlers tend to be more prosperous today than those largely bypassed by migration. The coefficient for external migrants is, however, larger than that for domestic migrants. The more than four-fold difference between both coefficients – statistically significant below the 1% level – hints at foreign migrants as a more powerful influence on the long-term prosperity of US counties than their American-born counterparts.

Regressions (3) to (5) shed light on the connection between different American-born population groups within a given county in 1880 and GDP per capita in 2010, using the share of external migrants as control variable. We first focus on the share of the population living in their state of birth in 1880 (regression 3) – including both individuals living in their county of birth and those who migrated within state borders. Counties with a large percentage of 'stayers' in 1880 had lower GDP per capita levels 130 years later (Regression 3): counties bypassed by migration – other than intra-state – more than a century ago were significantly poorer in 2010. Hence, a largely homogenous county population in 1880 has represented a serious barrier for long-term development.

Regressions (4) and (5) in Table 2 focus on domestic inter-state migration, distinguishing between migration from neighbouring states, on the one hand, and from non-adjacent states, on the other. The former is displayed in regression (4). No significant relation between the share of a county's population from neighbouring states and average income per capita in 2010 is found. This result may be a consequence of the similarities in institutional baggage, culture, traditions, and customs, between locals and migrants from adjacent states. Migrants from adjacent states may not be adding sufficient diversity to trigger additional economic dynamism.

Regression (5) dwells on interstate migration from non-adjacent states. The results point to a strong positive and significant relation between the current economic development of US counties and the presence of domestic migrants from more distant locations in 1880.

Compared to our main variables of interest in regressions (3) and (4), the significance level and the size of the coefficient are largest in (5). Domestic migrants travelling farther seem to have left a positive and enduring mark on the economic development of their receiving counties. A more heterogeneous and thus more diverse population composition in the receiving territory is associated with higher economic development. American-born migrants moving over long distances have thus left a similar economic imprint on local economic development to that of external migrants. The economic legacy of locals and migrants from neighbouring states is, in contrast, much more limited.

These results hint to the fact that the distance travelled by migrants has a considerable effect on long-term county economic growth. By dividing internal migrants into three groups – 'stayers' and within-state migrants, migrants from adjacent states, and long-distance migrants – domestic migrants have been implicitly grouped by distance travelled between birthplace and residence. The results indicate that the bigger the distance travelled, the greater the long-term economic legacy of internal migration.

Table 2. The long-term impact of internal migration on economic development – OLS 1880

Dep. Var.: income per capita 2010 (ln)	(1)	(2)	(3)	(4)	(5)
All migrants 1880	0.138***				
	(0.0251)				
Internal migrants 1880		0.0587**			
		(0.0282)			
Pop. same state 1880			-0.0587**		
			(0.0282)		
Pop. neighbour state 1880				-0.0489	
				(0.0360)	
Pop. rest of country 1880					0.134***
					(0.0388)
External migrants 1880		0.260***	0.260***	0.239***	0.279***
		(0.0401)	(0.0401)	(0.0400)	(0.0399)
Education 2000	0.0114***	0.0112***	0.0112***	0.0112***	0.0112***
	(0.000756)	(0.000750)	(0.000750)	(0.000751)	(0.000749)
Population 2000 (ln)	0.00388	0.00367	0.00367	0.00344	0.00340
	(0.00363)	(0.00366)	(0.00366)	(0.00367)	(0.00365)
Manufacturing 2000	-0.145***	-0.144***	-0.144***	-0.142***	-0.144***
	(0.0244)	(0.0242)	(0.0242)	(0.0242)	(0.0243)
Black population 2000	-0.00110***	-0.00101***	-0.00101***	-0.000950***	-0.00102***
	(0.000345)	(0.000340)	(0.000340)	(0.000339)	(0.000339)
Female 2000	0.0148***	0.0152***	0.0152***	0.0155***	0.0153***
	(0.00152)	(0.00153)	(0.00153)	(0.00153)	(0.00152)
Female part. 2000	0.00121	0.00112	0.00112	0.00106	0.00114
	(0.000786)	(0.000782)	(0.000782)	(0.000779)	(0.000781)
Unemployment 2000	-0.0203***	-0.0222***	-0.0222***	-0.0225***	-0.0220***
	(0.00320)	(0.00327)	(0.00327)	(0.00324)	(0.00327)
Infant mortality 2000	0.000201	0.000211	0.000211	0.000242	0.000214
	(0.000434)	(0.000444)	(0.000444)	(0.000449)	(0.000439)
Income 1880 (ln)	0.0133	0.0125	0.0125	0.0161	0.00295
	(0.0194)	(0.0192)	(0.0192)	(0.0190)	(0.0191)
Literacy 1880	0.133***	0.161***	0.161***	0.170***	0.142***
	(0.0293)	(0.0307)	(0.0307)	(0.0307)	(0.0311)
Population 1880 (ln)	-0.00996**	-0.0128***	-0.0128***	-0.0128***	-0.0137***
	(0.00440)	(0.00449)	(0.00449)	(0.00454)	(0.00447)
Manufacturing 1880	0.0105	-0.0189	-0.0189	-0.0175	-0.0114
	(0.0388)	(0.0374)	(0.0374)	(0.0375)	(0.0372)
Black population 1880	0.230***	0.241***	0.241***	0.227***	0.227***
	(0.0377)	(0.0378)	(0.0378)	(0.0378)	(0.0376)
Female 1880	0.101	0.113	0.113	0.0761	0.179
	(0.118)	(0.116)	(0.116)	(0.115)	(0.117)
Female part. 1880	-0.0667	-0.0621	-0.0621	-0.0462	-0.0482
	(0.0988)	(0.0974)	(0.0974)	(0.0964)	(0.0956)
Unemployment 1880	-0.00237	-0.00460	-0.00460	-0.00503	-0.00431
	(0.0115)	(0.0116)	(0.0116)	(0.0117)	(0.0115)
Observations	2,440	2,440	2,440	2,440	2,440
States	Yes	Yes	Yes	Yes	Yes
R-squared	0.682	0.686	0.686	0.686	0.688

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The significance levels and signs of the control variables reinforce the validity of the results. All variables included in the two vectors controlling for wealth influencing factors both at the time of migration and today show the expected signs. The lagged controls indicate that counties with a more educated, gender-balanced population, less employment in manufacturing, a smaller share of black inhabitants, and a lower unemployment rate were richer in 2010 than those with a less educated, largely male population, with a high share of black people, large employment in manufacturing, and a high unemployment rate.

Most 1880 control factors are insignificant, meaning that whatever influence they had on economic development a century ago has waned or disappeared altogether. The coefficient for the initial income level – one of the potentially strongest pull-factors at the time of migration – is insignificant. There are some exceptions, though. Literacy and the size of the black population of a county in 1880 are strongly positively connected to current levels of development. By contrast, the size of a county's population more than a century ago is associated with lower levels of development.

Distance travelled

In order to assess more accurately the extent to which the average distance covered by internal migrants affects the link between migration and long-term regional economic growth, Model 2 is run in its two different specifications focusing on, first, the average distance travelled by the entire migrant population – both foreign- and American-born – of a given county in 1880 and, second, on the average distance covered exclusively by domestic migrants crossing state-lines between their state of birth and destination. Table 3 reports the results for the OLS analysis – with the IV estimations for the base year 1880 included in Appendix 5.

Regression 1 shows that the average distance travelled by the entire migrant population in a county has a positive and highly significant impact on long-term economic development: the farther a migrant travelled before settling down, the stronger the impact. As this positive effect can be driven by the presence of international migrants having travelled for thousands of kilometres, in a second step, the analysis is performed focusing exclusively on American-born migrants crossing state lines. Regression 2 displays the results. Again, the link between the average distance travelled by internal migrants and the long-term economic performance of the receiving counties is positive and strongly significant (regression 2). Consequently, the

more a county managed to attract large shares of internal migrants from faraway places, the higher its income per capita 130 years later.

Table 3. The long-term impact of average migration distance on economic development, 1880 (OLS)

	OLS	
Dep. Var.: income per capita 2010 (ln)	(1) All Migrants	(2) Internal migrants
Distance all mig (ln) 1880	0.0376***	
	(0.00550)	
Distance int. mig (ln) 1880		0.0312***
		(0.0102)
Education 2000	0.0114***	0.0114***
	(0.000758)	(0.000764)
Population 2000 (ln)	0.00261	0.00324
	(0.00361)	(0.00364)
Manufacturing 2000	-0.137***	-0.139***
	(0.0246)	(0.0246)
Black population 2000	-0.000986***	-0.00101***
	(0.000343)	(0.000344)
Female 2000	0.0157***	0.0153***
	(0.00153)	(0.00153)
Female participation 2000	0.000834	0.00116
	(0.000778)	(0.000786)
Unemployment 2000	-0.0195***	-0.0197***
	(0.00312)	(0.00313)
Infant mortality 2000	0.000229	0.000268
	(0.000442)	(0.000436)
Income 1880 (ln)	0.0150	0.0226
	(0.0190)	(0.0195)
Literacy 1880	0.120***	0.137***
	(0.0292)	(0.0294)
Population 1880 (ln)	-0.00864*	-0.00669
	(0.00442)	(0.00451)
Manufacturing 1880	0.00203	0.0326
	(0.0373)	(0.0402)
Black population 1880	0.186***	0.192***
	(0.0377)	(0.0377)
Female 1880	-0.0468	-0.0277
	(0.112)	(0.115)
Female participation 1880	-0.0509	-0.0441
	(0.0975)	(0.0984)
Unemployment 1880	-0.00598	-0.00391
	(0.0116)	(0.0118)
Observations	2,444	2,441
States	Yes	Yes
R-squared	0.684	0.678
First stage F-stat	-	-

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The coefficients for both sets of controls are in line with those reported for the previous analysis.

Two robustness checks are undertaken to validate these results. Firstly, as in Model 1, we shift our base year by 30 years and run Model 2 based on 1910 migrants and average distances travelled. The results are displayed in Appendix 8 and present an almost carbon copy of those reported for 1880. Average distance travelled by the migrant population of any given US county in 1910 is in both cases highly significant and positively associated with income per capita levels in 2010. No matter which base year, long-distance migrants leave a greater economic trace than short-distance ones.

Secondly, to address potential endogeneity issues, an IV estimation is conducted. The results are displayed in Appendix 5 and reconfirm the findings of the OLS analysis.

In short, US domestic 19th century migrants have been highly beneficial for the long-term economic development of the receiving counties. A larger share of internal migrants more than 100 years ago has determined the economic fortunes of US counties in a way that is still evident today. Distance acts as crucial factor in this relationship. The larger the average distance travelled by the internal migrant, the longer-lasting the positive economic effect of migration. Simply crossing county-lines does not do the trick. The biggest benefits were reaped by counties which attracted migrants travelling over longer distances, precisely those who are more likely to bring a different institutional baggage from that of the local population. A more heterogeneous population, marked by the adventurous and entrepreneurial spirit of the long distance traveller seeking a new life in a faraway place, seems to have generated the seed of long-term economic development.

6. Conclusion

With international migration high on the priority list of current policy discussions, the majority of migration research has bypassed within-country migration. As a whole, the macroeconomic impact of internal migrants still remains poorly understood and the results – mainly focusing on the short-term impact – are far from conclusive. Moreover, the study of the long-term impact of internal migration on regional economic development as well as the analysis of the geographical distance covered by migrants have remained largely neglected. This paper has addressed these important gaps in the literature, examining the impact of

domestic US migration and the average distance covered by migrants in 1880 and 1910 on the wealth of US counties in 2010.

Two main research questions have driven the analysis: (1) Did late 19th century internal migrants transform the economy of their counties of destination in a way that can still be felt today?; and (2) Does the geographical distance covered by the immigrant matter for the impact of internal migration on long-term economic development?

The findings confirm that internal migrants having crossed state-lines between their birth state and destination exert a significant and positive long-term impact on the economic performance of the receiving regions. They leave a trace which is still evident more than 100 years after the settlement took place. Counties that attracted a large share of domestic migrants around the turn of the 20th century have become and remain more prosperous in 2010 than those largely bypassed by internal out-of-state migration streams. Similar to their foreign-born counterparts (Rodríguez-Pose and von Berlepsch, 2014), 19th century internal migrants have been one powerful force shaping the subsequent economic dynamism of US counties.

The geographical distance travelled by migrants has played a decisive role in this relationship. When analysing the average distance covered by internal migrants, the bigger the distance travelled, the greater the positive long-term economic legacy. Counties that drew a large number of long-distance migrants around the turn of the 20th century have been more dynamic over the next century. Large shares of population either born locally or in the same state are, in contrast, associated with significantly lower regional economic development over the long-term.

Over the past century, internal migrants – and especially those having crossed vast distances to arrive at their final destination – have acted as decisive force for regional economic growth in the US. Unfortunately, we can only make assumptions about the exact mechanisms as to why and how this is the case. Bringing their habits, customs and institutions from far-flung areas of the country to their receiving regions more than one hundred years ago increased regional diversity levels, creating more heterogeneous societies. The enterprising, hard-working, and risk-seeking spirit associated with the long-distance migrant brought novel ideas, experiences and abilities of the East to the once sparsely populated western counties. By contrast, places that did not manage to attract internal long-distance migrants – those that remained largely homogenous in their population composition – did not achieve a similar

economic dynamism as they lacked the enlarged skillset, the different ideas, experiences and abilities long-distance migrants brought along with them.

Internal long-distance migrants created a fertile ground for economic dynamism which lies at the root of a territorial prosperity that is much longer-lived than could have been imagined. Historic internal migrants thus represented much more than pure increases in the local supply of labour or a mere population redistribution at one point in time. They planted the seeds of a remarkably resilient long-term prosperity – a legacy which has determined the economic dynamism and vibrancy of places in the US for more than a century. More research, following case-study methodologies, will, nevertheless, be needed to properly disentangle the transfer mechanisms over time.

The research also highlights that ignoring the important legacy of worker mobility is risky. In a country made by migrants, short-sighted and short-term migration policies may stymie the possibility of important long-term economic impacts associated to facilitating the mobility of people looking for jobs and a better life for themselves and their children. Limiting this type of migration may consequently mean missing out on a huge economic potential and on a force for sustainable economic development which will be felt not just for years, but for decades and even centuries to come.

References

Acemoglu, D., Johnson, S., and Robinson, J. (2001): The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, 91(5), 1369–1401.

Alesina, A., and La Ferrara, E. (2005): Ethnic Diversity and Economic Performance. *Journal of Economic Literature*, 43(3), 762–800.

Åkerman, S. (1975): Internal migration, industrialisation and urbanisation (1895–1930): a summary of the Västmanland study. *Scandinavian Economic History Review*, 23(2), 149–158.

Akin, D., and Dökmeci, V. (2015): Cluster Analysis of Interregional Migration in Turkey. *Journal of Urban Planning and Development*, 141(3).

Arbia, G., Battisti, M. and Di Vaio, G. (2010): Institutions and geography: Empirical test of spatial growth models for European regions. *Economic Modelling* 27(1): 12-21.

Atack, J. and Passell, P. (1994): *A new economic view of American history: From colonial times to 1940*. P. Passell and S. Lee (Eds.). Second edition, New York: Norton.

Audretsch, D., Dohse, D. and Niebuhr, A. (2010): Cultural diversity and entrepreneurship: a regional analysis for Germany, *An International Journal of Urban, Regional and Environmental Research and Policy,* 45(1), 55–85.

Bell, M., Charles-Edwards, E., Kupiszewska, D., Kupiszewski, M., Stillwell, J., and Zhu, Y. (2015): Internal Migration Data Around the World: Assessing contemporary practice. *Population, Space and Place*, 21(1), 1–17.

Berker, A. (2011): Labor-Market Consequences of Internal Migration in Turkey. *Economic Development and Cultural Change*, 60(1), 197–239.

Biagi, B., Faggian, A., and McCann, P. (2011): Long and Short Distance Migration in Italy: The Role of Economic, Social and Environmental Characteristics. *Spatial Economic Analysis*, 6(1), 111–131.

Borjas, G., Bronars, S., and Trejo, S. (1992): Self-selection and internal migration in the United States. *Journal of Urban Economics*, 32(2), 159–185.

Borjas, G. (1995): The economic benefits from immigration. *Journal of Economic Perspectives*, 9(2), 3–22.

Card, D. (2005): Is the new immigration really so bad? *Economic Journal*, 115(507), 300–323.

Delisle, F., and Shearmur, R. (2010): Where does all the talent flow? Migration of young graduates and nongraduates, Canada 1996-2001. *Canadian Geographer-Géographe Canadien*, 54(3), 305–323.

Duranton, G., Rodríguez-Pose, A and Sandall, R (2009): Family Types and the Persistence of Regional Disparities in Europe. *Economic Geography,* 85(1), 23–47.

Ellis, M. (2012): Reinventing US Internal Migration Studies in the Age of International Migration. *Population, Space and Place*, 18(2), 196–208.

Ferrie, J. (2005): History Lessons: The End of American Exceptionalism? Mobility in the United States Since 1850. *Journal of Economic Perspectives*, 19(3), 199–215.

Florida, R. (2002): *The rise of the creative class: and how it's transforming work, leisure, community and everyday life*. New York, NY: Basic Books.

Gallman, R. (2000): Economic Growth and Structural Change in the Long Nineteenth Century. In R. Engerman and R. Gallman (Eds.), *The Cambridge Economic History of the United States: The long Nineteenth Century,* 1–56. Cambridge: Cambridge University Press.

Gill, I., and Raiser, M. (2012): *Golden Growth: Restoring the lustre of the European economic model*. The World Bank, Washington, DC. Retrieved 14.10.2017 under http://documents.worldbank.org/curated/en/539371468036253854/Main-report

Gordon, M. (1954). *Employment expansion and population growth, the California experience: 1900-1950.* Berkeley: University of California Press.

Gordon, I. and McCann, P. (2005): Innovation, agglomeration, and regional development. *Journal of Economic Geography*, 5, 523–543.

Greenwood, M. (1997): Chapter 12 Internal migration in developed countries. *Handbook of Population and Family Economics*, 1, 647–720.

Greenwood, M. (2015): Internal Migration in Industrialized Countries. In *International Encyclopaedia of the Social and Behavioral Sciences,* 443–446.

Gregory, J. (2005). *The Southern Diaspora: How the Great Migrations of Black and White Southerners Transformed America*. Chapel Hill: University of North Carolina Press.

Haines, M. (2000). The Population of the United States, 1790–1920. In S. Engerman and R. Gallman (Eds.), *The Cambridge Economic History of the United States: The long Nineteenth Century,* 143–206. Cambridge: Cambridge University Press.

Hipp, J. and Boessen, A. (2016): The Shape of Mobility: Measuring the Distance Decay Function of Household Mobility. *The Professional Geographer*, 1–13.

Huning, S., Bens, O., and Huettl, R. (2012): Demographic Change beyond the Urban-Rural Divide: Re-Framing Spatial Differentiation in the Context of Migration Flows and Social Networks. *Erde*, 143(1–2), 153–172.

Jacobs, J. (1969): *The economy of cities*. New York, NY: Random House.

Joly, D. (2000): Some structural effects of migration on receiving and sending countries. *International Migration*, 38(5), 25–40.

Kemeny, T., and Cooke, A. (2017): Urban immigrant diversity and inclusive institutions. *Economic Geography*, 93(3), 267-291.

Kuhn, R. (2015): Internal Migration: Developing Countries. In *International Encyclopaedia of the Social and Behavioral Sciences,* 433–442.

Kupiszewska, D. and Nowok, B. (2008): Comparability of statistics on international migration flows in the European Union. In J. Raymer and F. Willekens (Eds.), *International Migration in Europe: Data, Models and Estimates*, 41–71. Chichester: Chichester.

Lee, E. (1966). A theory of migration. *Demography*, *3*(1), 47–57.

Lundborg, P. and Segerström, P. (2002): The growth and welfare effects of international mass migration. *Journal of International Economics*, 56, 177–204.

Merk, F. (1978): *History of the Westward Movement*. New York: Alfred A. Knopf.

Molloy, R., Smith, C. and Wozniak, A. (2011): Internal Migration in the United States. *Journal of Economic Perspectives*, 25(3), 173–196.

Newbold, K. and Bell, M. (2001): Return and Onwards Migration in Canada and Australia: Evidence from Fixed Interval Data. *International Migration Review*, 35(4), 1157–1184.

Niedomysl, T. and Fransson, U. (2014): On Distance and the Spatial Dimension in the Definition of Internal Migration. *Annals of the Association of American Geographers*, 104(2), 357–372.

Niedomysl, T., Ernstson, U. and Fransson, U. (2017): The Accuracy of Migration Distance Measures. *Population, Space and Place*, 23(1).

Niedomysl, T. (2011): How Migration Motives Change over Migration Distance: Evidence on Variation across Socio-economic and Demographic Groups. *Regional Studies*, 45(6), 843–855.

Ortega, F., and Peri, G. (2009): The Causes and Effects of International Migrations: Evidence from OECD Countries 1980-2005. *NBER Working Paper Series*, No. 14833.

Ottaviano, G. and Peri, G. (2006): The economic value of cultural diversity: evidence from US cities. *Journal of Economic Geography*, 6, 9–44.

Özgen, C., Nijkamp, P. and Poot, J. (2011): Immigration and innovation in European regions. *IZA Discussion Papers*, Institute for the Study of Labor (IZA), No. 5676.

Partridge, J. and Furtan, H. (2008): Increasing Canada's international competitiveness: Is there a link between skilled immigrants and innovation? Paper presented at the *American Agricultural Economics Association annual meeting*, Orlando, FL.

Pendakur, R. and Young, N. (2013): Putting on the moves: Individual, household, and community-level determinants of residential mobility in Canada. *Demographic research*, 29, 767–795.

Petersen, W. (1986): *Dictionary of demography: terms, concepts and institutions*. R. Petersen (Ed.). New York: Greenwood Press.

Pierson, G. (2011): The moving American. *Yale Review*, 99(3), 90–103.

Plane, D. (1993): Demographic Influences on Migration. *Regional Studies*, 27(4), 375–383.

Plane, D. and Rogerson, P. (2015): On Tracking and Disaggregating Center Points of Population. *Annals of the Association of American Geographers,* 105(5), 968–986.

Poot, J. (2008): Demographic change and regional competitiveness: the effects of immigration and ageing. *International Journal of Foresight and Innovation Policy*, 4(1/2), 129–145.

Putnam, R. (1993): *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.

Ravenstein, E. (1885). The Laws of Migration. *Journal of the Statistical Society of London*, *48*(2), 167–235.

Ravenstein, E. (1889). The Laws of Migration. *Journal of the Royal Statistical Society*, *52*(2), 241–305.

Rees, P., Bell, M., Duke-Williams, O. and Blake, M. (2000). Problems and solutions in the measurement of migration intensities: Australia and Britain compared. *Population Studies*, 54(2), 207–222.

Rodríguez-Pose, A. and von Berlepsch, V. (2014): The legacy of mass migration on economic development in the United States. *Annals of the Association of American Geographers*, 104(3), 628–651.

Rodríguez-Pose, A. and von Berlepsch, V. (2015): European migration, national origin and long-term economic development in the United States. *Economic Geography*, 91(4), 393–424.

Rodríguez-Pose, A. and von Berlepsch, V. (2018): Does population diversity matter for economic development in the very long term? Historic migration, diversity and county wealth in the US. *European Journal of Population*. https://doi.org/10.1007/s10680-018-9507-z

Rogers, A., Raymer, J. and Newbold, K. (2003): Reconciling and translating migration data collected over time intervals of differing widths. *The Annals of Regional Science*, 37(4), 581–601.

Ruggles, S., Genadek, K., Goeken, R., Grover, J. and Sobek, M. (2015): *Integrated Public Use Microdata Series: Version 6.0*. Minneapolis: University of Minnesota.

Rupasingha, A., Liu, Y. and Partridge, M. (2015): Rural Bound: Determinants of Metro to Non-Metro Migration in the United States. *American Journal of Agricultural Economics*, 97(3), 680–700.

Saxenian, A. (2006): *The new Argonauts: regional advantage in a global economy*. Cambridge, MA: Harvard University Press.

Skeldon, R. (2006): Interlinkages between internal and international migration and development in the Asian region. *Population, Space and Place*, 12(1), 15–30.

Sloan, M. and Morrison, P. (2016): The post-move satisfaction of internal migrants. *Migration Studies*, 4(1), 97–115.

Spickard, P. (2007): *Almost all aliens: immigration, race, and colonialism in American history and identity*. New York: Routledge.

Staiger, D. and Stock, J. (1997): Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3), 557–586.

Stock, J. and M. Yogo (2005): Testing for weak instruments in linear IV regression. In: Donald, A. and Stock, J. (Eds.), *Identification and Inference for Econometric Models*. New York, NY: Cambridge University Press.

Tabellini, G. (2010): Culture and institutions: Economic development in the regions of Europe. *Journal of the European Economic Association*, 8(4), 677–716.

Thomas, D. (1936): Internal Migrations in Sweden: A Note on Their Extensiveness as Compared with Net Migration Gain or Loss. *American Journal of Sociology*, 42(3), 345–357.

Thomas, D. (1941): *Social and economic aspects of Swedish population movements, 1750-1933*. New York: The Macmillan Company.

Thomas, D. and Kuznets, S. (1957): *Population redistribution and economic growth: United States, 1870-1950*. Philadelphia: the American Philosophical Society.

Tolnay, S., White, K., Crowder, K., and Adelman, R. (2005). Distances Traveled during the Great Migration: An Analysis of Racial Differences among Male Migrants. *Social Science History*, *29*(4), 523–548.

United Nations (2016): *International Migration Report 2015*. Department of Economic and Social Affairs, United Nations, New York.

U.S. Census Bureau (2016): *Annual Geographical Mobility Rates, by type of movement: 1948-2016*. US Census Bureau, Current Population Survey. Retrieved 09.02.2017 under www.census.gov/data/tables/time-series/demo/geographic-mobility/historic.html

U.S. Bureau of the Census (1972): *U.S. Census of Population 1970*. Department of Commerce, Bureau of the Census, Government Printing Office, Washington, D.C.

von Berlepsch, V., Rodríguez-Pose, A. and Lee, N. (2018) A woman's touch? Female migration and economic development in the United States. *Regional Studies* 53(1), 131-145.

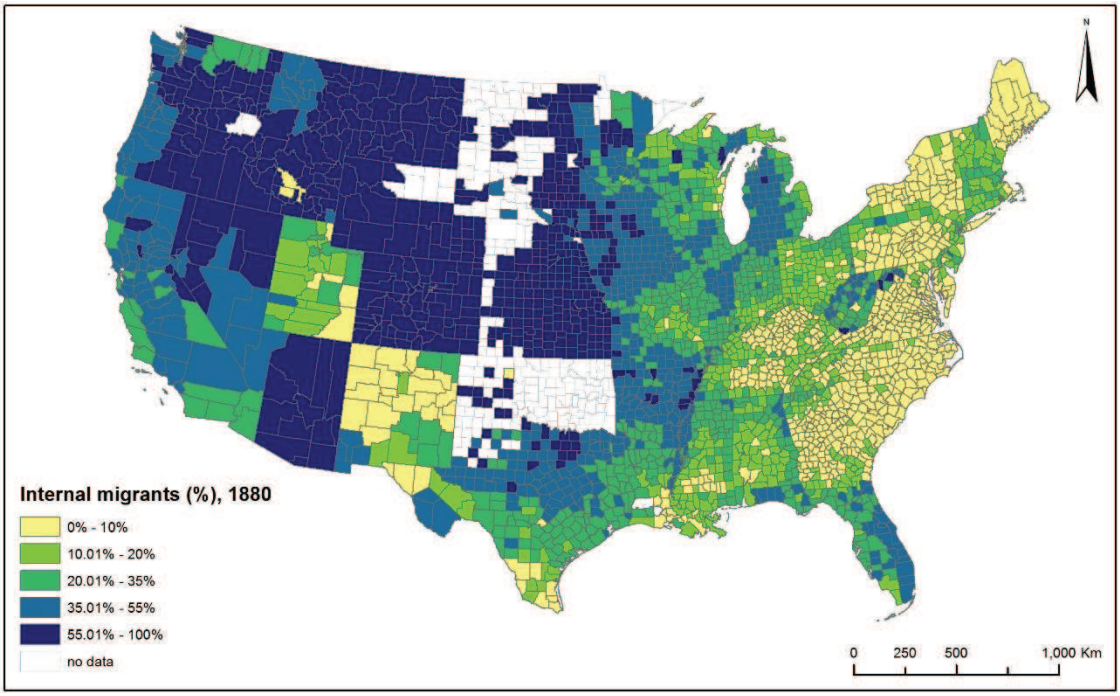
White, M. and Lindstrom, D. (2005): Chapter 11: Internal Migration. In *Handbook of Population*, 311–346, Springer Science and Business Media B.V.

Wilkerson, I. (2010). *The Warmth of Other Suns: The Epic Story of America's Great Migration*. New York: Random House.

Yap, L. (1976): Internal Migration and Economic Development in Brazil. *The Quarterly Journal of Economics*, 90(1), 119–137.

Appendix 1

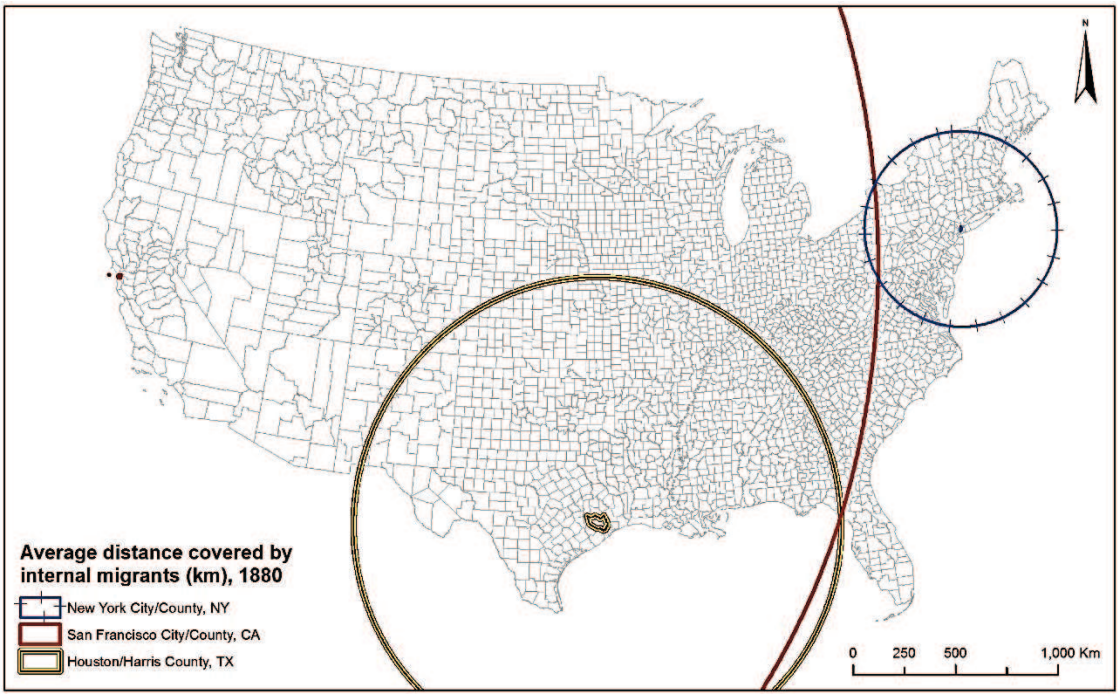
American-born internal migrants by county, 1880 (% of total American-born population)



Source: Ruggles, et al. 2015; own calculations

Appendix 2

Average migrant distance radius from Houston, New York, and San Francisco, 1880



Source: Ruggles, et al. 2015; own calculations

Appendix 3

Descriptive statistics dependent variable and main independent variables of interest

Variable	Obs.	Mean	Std. dev.	Min	Max
Income per capita 2010	3142	34072,8	7764,75	16023	110956
All migrants 1880	2858	0,37	0,.2770218	0,002884	1
Internal migrants 1880	2921	0,33	0,2653268	0	1
Population same state 1880	2921	0,67	0,2653268	0	1
Population neighbour state 1880	2921	0,10	0,1005804	0	0,8552632
Population rest of country 1880	2921	0,23	0,2465148	0	1
Av. distance int. migrants 1880 (km)	2921	898,51	669,41	0	3855,5
Av. distance all migrants 1880 (km)	2921	2.140,74	1.527,82	0	8875,2
All migrants 1910	3131	0,33	0,2560091	0,000223	1
Internal migrants 1910	3135	0,26	0,2274818	0	1
Population same state 1910	3135	0,74	0,2274818	0	1
Population neighbour state 1910	3135	0,11	0,1028442	0	0,88
Population rest of country 1910	3135	0,16	0,1847167	0	1
Av. distance int. migrants 1910 (km)	3135	780, 45	560,64	0	3362,7
Av. distance all migrants 1910 (km)	3135	2.138,44	1.710,03	0	9208,0

Descriptive statistics control variables

Variable	Obs.	Mean	Std. dev.	Min	Max
Education 2000	3143	0,17	0,08	0,05	0,64
Population 2000	3144	89920,43	292369,90	67	9519338
Black population 2000	3120	0,09	0,145282	0	0,87
Female 2000	3144	0,50	0,019063	0,33	0,57
Female participation 2000	3142	0,55	0,065496	0,27	0,81
Unemployment 2000	3141	0,04	0,016573	0,01	0,18
Infant mortality 2000	3142	7	7,495610	0	93,8
Manufacturing 2000	3049	0,14	0,104576	0,003	1
Mean income 1880	2877	6,36	2,872315	1,75	30,27
Literacy 1880	2921	0,74	0,227872	0	1
Population 1880	3045	17799,95	39283,61	0	1206299
Black population 1880	2994	0,14	0,215203	0	0,92
Female 1880	2994	0,45	0,091157	0	0,88
Female participation 1880	2921	0,06	0,060511	0	0,39
Unemployment 1880	2918	0,16	0,295097	0	5
Manufacturing 1880	2921	0,06	0,085532	0	0,91
Mean income 1910	3128	6,82	2,132763	0	21,74
Literacy 1910	3125	0,89	0,1183161	0,333333	1
Population 1910	3138	31277,43	103795,6	0	2762522
Black population 1910	3135	0,02	0,0827582	0	0,717431
Female 1910	3135	0,47	0,0331056	0,278877	0,553718
Female participation 1910	3125	0,10	0,0721319	0	0,449153
Unemployment 1910	3128	0,06	0,0346171	0	0,484472
Manufacturing 1910	3122	0,097	0,115119	0	0,75

Appendix 4

Explanation to variables and sources

Variable	Description	Source
Main variables of interest		
Inc per cap 2010 (ln)	Natural log of average income per capita levels in current US dollars at county level for county *i* in year *t* = 2010 (unadjusted for inflation)	US BEA
All migrants~	Share of total migrants, both foreign-born and American-born, relative to total population in county *i* in year t_0	IPUMS USA
Internal migrants~	Share of American-born internal migrants relative to total population in county *i* in year t_0 (internal migrants refer to all current residents having crossed state-lines between birth state and current state of residence)	IPUMS USA
Population same state~	Fraction of American-born population in county *i* in year t_0 relative to total American-born population with birth state equivalent to state of residence	IPUMS USA
Population neighbour state~	Fraction of American-born population in county *i* in year t_0 relative to total American-born population with birth state equivalent to any adjacent state of state of residence	IPUMS USA
Population rest of country~	Fraction of American-born population in county *i* in year t_0 relative to total American-born population with birth state equivalent to any non-adjacent state of state of residence	IPUMS USA
External migrants~	Share of foreign-born migrants relative to total population in county *i* in year t_0	IPUMS USA
Distance all mig (ln)~	Natural log of average distance travelled between birth state/birth country and county of residence by all migrants, both foreign-born and American-born, living in county *i* in year t_0	IPUMS USA
Distance int. mig (ln)~	Natural log of average distance travelled between birth state and county of residence by American-born internal migrants living in county *i* in year t_0	IPUMS USA

Variable	Description	Source
Instruments		
Topography	Scale variable extracted from the National Atlas of the United States of America of the U.S. Geological Survey (1970) published as part of the Natural Amenities Scale corresponding to topography type of county *i* with range [1-21]	U.S. Department of Interior
Water area	Natural log of size of water areas within a county measured in square miles retrieved from the TIGER Geodatabases	U.S. Census Geography division
Control Variables		
Education	Percentage of population of county *i* with college degree in *t-k*	ICPSR
Literacy~	Literacy rate in county *i* in t_0	IPUMS USA
Income~	Constructed mean income score on county level. Individual income levels assigned to occupational data on the basis of median total income per occupation in hundreds of 1950 dollars, as natural log in t_0	IPUMS USA
Population (ln)~	Natural log of total population of county *i* in t_0 and *t-k*	ICPSR
Manufacturing~	Percentage of labour force employed in manufacturing in county *i* in t_0 and *t-k*	US BLS and IPUMS USA
Black Population~	Percentage of black population in county *i* in t_0 and *t-k*	ICPSR
Female~	Percentage of female population in county *i* in t_0 and *t-k*	Census 2000 summary files and IPUMS USA
Female Participation~	Female participation rate in the labour force in county *i* in t_0 and *t-k*	ICPSR and IPUMS USA
Unemployment~	Unemployment rate in county *i* in t_0 and *t-k* Proxy in 1880 (unemployment rate not available): Months unemployed last year	IPUMS USA and US BLS
Infant mortality~	Infant mortality rate in county *i* in *t-k* measured as number of deaths among infants aged <1 year per 1,000 live births	CDC
State Controls	State dummies	Own construction

~ refers to respective year in question: 1880/1910 for variables of interest and 1880/1910/2000 for control variables

Appendix 5

Dealing with endogeneity: Instrumental Variable estimation

Any analysis of long term migration data is prone to endogeneity issues. Potential spatial sorting, omitted variables, or reverse causality issues can pose threats to the internal validity and hence distort retrieved OLS results. In order to identify the underlying link between internal migration and economic development, possible endogeneity biases in the least-squares estimates are addressed using an instrumental variable (IV) estimation. For both models, geography serves as an exogenous source of variation. Two instruments are selected: Topography – as instrument for internal migration (Model 1) – and size of the water area in a county – instrumenting the average distance travelled (Model 2). Both instruments satisfy the relevance criterion of IV analysis, as both retrieve sufficiently large first-stage F-statistics based on the Staiger and Stock (1997) test for weak instruments in combination with the Stock and Yogo (2005) critical values.

Table A5.1. Description of topography scale retrieved from The National Atlas of the USA Land Surface Topography Codes

Category	Code	Land surface
Plains	1	Flat plains
	2	Smooth plains
	3	Irregular plains, slight relief
	4	Irregular plains
Tablelands	5	Tablelands, moderate relief
	6	Tablelands, considerable relief
	7	Tablelands, high relief
	8	Tablelands, very high relief
Plains with Hills or Mountains	9	Plains with hills
	10	Plains with high hills
	11	Plains with low mountains
	12	Plains with high mountains
Open Hills and Mountains	13	Open low hills
	14	Open hills
	15	Open high hills
	16	Open low mountains
Hills and Mountains	18	Hills
	19	High hills
	20	Low mountains
	21	High mountains

Source: U.S. Department of Interior, U.S. Geological Survey, Washington, DC., 1970

In Model 1, topography is instrumented for internal migrants. The topography variable is extracted from the National Atlas of the USA (US Geological Survey), published as part of the US Natural Amenities Scale of the US Department of Agriculture (Table A5.1). The 21 level-scale categorises land surface forms at county-level ranging from flat plains and tablelands to hillsides and mountains. Topography is exogenous to income per capita in 2010 – and hence highly likely to be fully uncorrelated with the error term. We assume that topography affects economic development at county level via the settlement pattern of internal migrants. Topography was a crucial factor behind domestic migrant settlement patterns over 100 years ago. The interaction of topography with climate and vegetation affects soil quality and served as a crucial pull-factor, raising incentives for migrants to pack up and move. Fresh good quality soils "[drew] migrants from worn-out lands in the East to virgin lands in the West" (Merk, 1978: 229). Cattle farmers, hunters, trappers, and miners were attracted by the vast lands of plateaus and hill sides. Grain-, pork farmers, and wool producers settled in the plains, while fur traders saw high potential in the plateaus and basins of the North West. The Gulf plains to the south attracted internal migrants focusing on cotton and sugar production (Merk, 1978).

The variation in the average distance travelled by migrants (Model 2) is instrumented using the size of water areas within a county. The data, measured in square miles, were retrieved from the TIGER Geodatabases by the US Census Geography division. Water areas, again exogenous to income per capita levels in 2010, were a decisive element for the migrant settlement pattern in the 19th century as well as decisive in determining the distance travelled by migrants. Navigable rivers provided important infrastructure and vital access to drinking water and irrigation of nearby lands. The better the waterways and water supply, the farther the migrant travelled. Waterways shaped long-distance migrant routes which, in turn, gave rise to migrant settlements along the way. Mountain ranges or rough landscapes such as the Appalachians or the Rocky Mountains could only be crossed following rivers, cutting deep canyons and valleys into the abrupt terrain. The main migrant trails across the country, such as the Oregon Trail in the North West crossing the northern part of the Rockies, followed tightly knit river systems. Even when the railroad network had grown considerably in size and efficiency, waterways remained a fundamental element in the country's infrastructure for both transportation as well as internal migrant settlement (Merk, 1978). Waterways facilitated

the journey, enlarged the distances travelled, and made quick progress possible. A lack of water, however, shortened migration routes considerably.

Model 1

The OLS results of the analysis of Model 1 presented in Table 2 could be affected by endogeneity issues caused by reverse causality, sorting, or omitted variable bias. In order to address these issues, we conduct an IV estimation, using, as indicated above, topography as instrument for the five different migrant population subgroups in a given US county in 1880. Table A5.2 displays the results.[5]

When using topography as IV to retrieve the underlying effect of past internal migration flows on income per capita levels more than 100 years later, the large majority of the OLS results are validated. Internal as well as external migration are confirmed as growth enhancing factors over the very long-term – either taken together as in regression (1) or apart as displayed in regression (2) of Table 2. Consequently, counties which attracted large inflows of both foreign and American-born migrants at the end of the 19th century display significantly larger income per capita levels in 2010 than those that lacked a strong migrant pool at the time. Large numbers of migrants, regardless of origin, have become an essential factor behind county-level economic development for more than a century.

The results of the OLS analysis are, once again, largely validated when focusing on the different American-born population groups in regressions (3) to (5) in Table A5.2. Counties bypassed by out-of-state domestic migration more than a century ago are today poorer than those which attracted migrants in large numbers (regression 3). Having a high percentage of locals and therefore low population diversity has been detrimental for long-term economic growth. By contrast, any type of cross-border internal migration has resulted in greater long-term economic dynamism and wealth at county level. This applies for both migrants from neighbouring states (regression 4) and from farther away (regression 5). The sway of domestic migrants crossing at least two state-lines in their journey to a better future remains in any case considerably higher, as indicated by the coefficient in regression 5. Distance thus largely explains the relation between internal migrants and long-term economic performance. The larger the distance travelled by internal migrants, the stronger their long-term impact on county income per capita levels.

[5] The following discussion of the results is equally valid for the IV analysis shifting our base year to 1910. Similar results for 1910 to those displayed in Table 3 can be found in Appendix 7.

Table A5.2. The long-term impact of internal migration on economic development – IV 1880

Dep. Var.: income per capita 2010 (ln)	(1)	(2)	(3)	(4)	(5)
All migrants 1880	1.123***				
	(0.258)				
Internal migrants 1880		1.497***			
		(0.433)			
Pop. same state 1880			-1.497***		
			(0.433)		
Pop. neighbour state 1880				2.299***	
				(0.787)	
Pop. rest of country 1880					4.291*
					(2.297)
External migrants 1880		0.730***	0.730***	0.341***	1.456**
		(0.151)	(0.151)	(0.0743)	(0.647)
Education 2000	0.0109***	0.0114***	0.0114***	0.0121***	0.0102***
	(0.000820)	(0.000895)	(0.000895)	(0.00102)	(0.00154)
Population 2000 (ln)	0.00550	0.00622	0.00622	0.0107	-0.00214
	(0.00493)	(0.00520)	(0.00520)	(0.00695)	(0.0114)
Manufacturing 2000	-0.163***	-0.169***	-0.169***	-0.168***	-0.170**
	(0.0346)	(0.0393)	(0.0393)	(0.0441)	(0.0750)
Black population 2000	-0.00170***	-0.00212***	-0.00212***	-0.00173***	-0.00284**
	(0.000472)	(0.000589)	(0.000589)	(0.000578)	(0.00125)
Female 2000	0.0128***	0.0110***	0.0110***	0.00990***	0.0131***
	(0.00192)	(0.00238)	(0.00238)	(0.00286)	(0.00409)
Female part. 2000	0.00172	0.00209*	0.00209*	0.00145	0.00330
	(0.00108)	(0.00121)	(0.00121)	(0.00127)	(0.00247)
Unemployment 2000	-0.0231***	-0.0162***	-0.0162***	-0.0197***	-0.00985
	(0.00525)	(0.00569)	(0.00569)	(0.00507)	(0.0115)
Infant mortality 2000	-0.000213	-0.000283	-0.000283	-0.000255	-0.000335
	(0.000601)	(0.000654)	(0.000654)	(0.000752)	(0.00138)
Income 1880 (ln)	-0.0995**	-0.104**	-0.104**	0.0760	-0.441*
	(0.0401)	(0.0463)	(0.0463)	(0.0552)	(0.258)
Literacy 1880	0.0225	-0.0934	-0.0934	0.276***	-0.784
	(0.0603)	(0.0942)	(0.0942)	(0.0729)	(0.542)
Population 1880 (ln)	-0.0285***	-0.0188***	-0.0188***	-0.00272	-0.0487*
	(0.00774)	(0.00687)	(0.00687)	(0.0106)	(0.0256)
Manufacturing 1880	-0.103	0.000230	0.000230	-0.132	0.247
	(0.0700)	(0.0662)	(0.0662)	(0.0849)	(0.193)
Black population 1880	0.417***	0.390***	0.390***	0.596***	0.00586
	(0.0672)	(0.0675)	(0.0675)	(0.130)	(0.172)
Female 1880	1.106***	1.135***	1.135***	-0.128	3.493*
	(0.321)	(0.370)	(0.370)	(0.251)	(1.888)
Female part. 1880	-0.235*	-0.258*	-0.258*	-0.473**	0.145
	(0.134)	(0.154)	(0.154)	(0.237)	(0.319)
Unemployment 1880	-0.00383	0.00460	0.00460	-0.00180	0.0165
	(0.0167)	(0.0184)	(0.0184)	(0.0192)	(0.0426)
Observations	2,435	2,435	2,435	2,435	2,435
States	Yes	Yes	Yes	Yes	Yes
First stage F-stat	36.29	20.45	20.45	15.00	8.16

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Model 2

Table A5.3. The long-term impact of average migration distance on economic development, 1880 (IV)

Dep. Var.: income per capita 2010 (ln)	IV	
	(3) All migrants	(4) Internal migrants
Distance all mig (ln) 1880	0.199***	
	(0.0509)	
Distance int. mig (ln) 1880		0.485***
		(0.122)
Education 2000	0.0114***	0.0102***
	(0.000790)	(0.000946)
Population 2000 (ln)	-0.00346	-0.00745
	(0.00447)	(0.00528)
Manufacturing 2000	-0.112***	-0.0910**
	(0.0332)	(0.0389)
Black population 2000	-0.000748*	-0.000841*
	(0.000423)	(0.000489)
Female 2000	0.0176***	0.0171***
	(0.00192)	(0.00206)
Female participation 2000	-0.000292	0.00220**
	(0.000929)	(0.00104)
Unemployment 2000	-0.0178***	-0.0166***
	(0.00375)	(0.00440)
Infant mortality 2000	0.000115	0.000478
	(0.000538)	(0.000546)
Income 1880 (ln)	-0.0467	-0.0788**
	(0.0285)	(0.0367)
Literacy 1880	0.0108	0.0111
	(0.0487)	(0.0547)
Population 1880 (ln)	-0.0127***	0.00555
	(0.00488)	(0.00633)
Manufacturing 1880	-0.105**	0.116**
	(0.0487)	(0.0560)
Black population 1880	0.107**	0.0257
	(0.0517)	(0.0686)
Female 1880	-0.0698	0.164
	(0.113)	(0.135)
Female participation 1880	-0.0935	-0.0130
	(0.111)	(0.130)
Unemployment 1880	-0.0144	-0.0192
	(0.0143)	(0.0161)
Observations	2,444	2,441
States	Yes	Yes
R-squared	-	-
First stage F-stat	24.35	28.27

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

To control for endogeneity in Model 2, we use the size of water areas in a given county as instrument for the average distance travelled by migrants. The results, displayed in regressions (3) and (4) of Table A5.3, reconfirm the findings of the OLS analysis: the average distance travelled by the migrant population of any given county in 1880 between birthplace and current residence has a positive impact on long-term economic growth. Both the average distance travelled by the total migrant population as well as by the internal migrant subgroup display positive and highly significant coefficients. The presence of a higher share of long-distance domestic migrants in 1880 has therefore been strongly beneficial for long-term economic growth at the county level.

Appendix 6

As a robustness test of the results presented in Table 2, we first shift the base migration year by 30 years to 1910, estimating again Model 1 including all five specifications of the independent variable of interest. The results are displayed in Table A6.1 below. Apart from slightly weaker significance levels, the coefficients for the different groups of internal migrants largely coincide with those obtained when using 1880 as base year. Both a large share of internal as well as external migrants – taken together in regression 1 and apart in regression 2 – is significantly associated with higher levels of economic development one century later. A large percentage of Americans still living in their state of birth is negatively connected with long-term growth (regression 3). The coefficient for the relationship between income per capita levels in 2010 and the share of internal migrants from neighbouring states in 1910 remains insignificant (regression 4). Again, the largest and most significant coefficient of the internal migrant subgroups pertains to the group of internal migrants from non-adjacent states (5).

The long-term impact of internal migration on economic development – OLS 1910

Dep. Var.: income per capita 2010 (ln)	(1)	(2)	(3)	(4)	(5)
All migrants 1910	0.0721***				
	(0.0220)				
Internal migrants 1910		0.0375*			
		(0.0215)			
Pop. same state 1910			-0.0375*		
			(0.0215)		
Pop. neighbour state 1910				-0.0149	
				(0.0337)	
Pop. rest of country 1910					0.0768**
					(0.0327)
External migrants 1910		0.198***	0.198***	0.186***	0.205***
		(0.0494)	(0.0494)	(0.0495)	(0.0493)
Education 2000	0.0116***	0.0114***	0.0114***	0.0114***	0.0114***
	(0.000759)	(0.000749)	(0.000749)	(0.000749)	(0.000747)
Population 2000 (ln)	-0.000998	-0.000598	-0.000598	0.000321	-0.00113
	(0.00468)	(0.00468)	(0.00468)	(0.00461)	(0.00465)
Manufacturing 2000	-0.130***	-0.131***	-0.131***	-0.131***	-0.131***
	(0.0242)	(0.0241)	(0.0241)	(0.0240)	(0.0240)
Black population 2000	0.000196	0.000278	0.000278	0.000299	0.000261
	(0.000273)	(0.000273)	(0.000273)	(0.000273)	(0.000273)
Female 2000	0.0163***	0.0165***	0.0165***	0.0168***	0.0166***
	(0.00152)	(0.00152)	(0.00152)	(0.00152)	(0.00152)
Female part. 2000	0.00214***	0.00218***	0.00218***	0.00208***	0.00217***
	(0.000761)	(0.000763)	(0.000763)	(0.000764)	(0.000761)
Unemployment 2000	-0.0221***	-0.0223***	-0.0223***	-0.0222***	-0.0225***
	(0.00297)	(0.00302)	(0.00302)	(0.00303)	(0.00299)
Infant mortality 2000	8.43e-05	7.11e-05	7.11e-05	6.59e-05	7.68e-05
	(0.000419)	(0.000418)	(0.000418)	(0.000421)	(0.000418)
Income 1910 (ln)	-0.000805	7.27e-05	7.27e-05	0.00507	-0.00441
	(0.0143)	(0.0143)	(0.0143)	(0.0137)	(0.0142)
Literacy 1910	0.123***	0.162***	0.162***	0.175***	0.157***
	(0.0388)	(0.0397)	(0.0397)	(0.0396)	(0.0398)
Population 1910 (ln)	7.61e-05	-0.00291	-0.00291	-0.00325	-0.00273
	(0.00587)	(0.00579)	(0.00579)	(0.00576)	(0.00579)
Manufacturing 1910	-0.0778***	-0.0918***	-0.0918***	-0.0929***	-0.0855***
	(0.0263)	(0.0267)	(0.0267)	(0.0267)	(0.0270)
Black population 1910	0.127**	0.134**	0.134**	0.135**	0.135**
	(0.0580)	(0.0582)	(0.0582)	(0.0583)	(0.0581)
Female 1910	-0.609***	-0.513***	-0.513***	-0.570***	-0.500***
	(0.157)	(0.164)	(0.164)	(0.165)	(0.162)
Female part. 1910	0.122**	0.122**	0.122**	0.119**	0.125**
	(0.0529)	(0.0527)	(0.0527)	(0.0526)	(0.0526)
Unemployment 1910	0.0598	0.0515	0.0515	0.0560	0.0544
	(0.0838)	(0.0837)	(0.0837)	(0.0837)	(0.0835)
Observations	2,617	2,617	2,617	2,617	2,617
States	Yes	Yes	Yes	Yes	Yes
R-squared	0.673	0.674	0.674	0.674	0.675

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix 7

The long-term impact of internal migration on economic development – IV 1910

Dep. Var.: income per capita 2010 (ln)	(1)	(2)	(3)	(4)	(5)
All migrants 1910	1.420***				
	(0.331)				
Internal migrants 1910		1.653***			
		(0.421)			
Pop. same state 1910			-1.653***		
			(0.421)		
Pop. neighbour state 1910				3.042***	
				(0.965)	
Pop. rest of country 1910					3.622**
					(1.428)
External migrants 1910		0.661***	0.661***	0.374***	1.003***
		(0.142)	(0.142)	(0.105)	(0.343)
Education 2000	0.0120***	0.0131***	0.0131***	0.0123***	0.0140***
	(0.000898)	(0.00107)	(0.00107)	(0.00113)	(0.00166)
Population 2000 (ln)	-0.0341***	-0.0371***	-0.0371***	-0.0136	-0.0651**
	(0.0107)	(0.0122)	(0.0122)	(0.0114)	(0.0278)
Manufacturing 2000	-0.147***	-0.136***	-0.136***	-0.136**	-0.135***
	(0.0384)	(0.0410)	(0.0410)	(0.0538)	(0.0525)
Black population 2000	-5.67e-05	-0.000573	-0.000573	0.000128	-0.00141
	(0.000484)	(0.000533)	(0.000533)	(0.000716)	(0.000867)
Female 2000	0.00910***	0.00726**	0.00726**	0.00699	0.00759
	(0.00306)	(0.00358)	(0.00358)	(0.00443)	(0.00512)
Female part. 2000	0.00596***	0.00575***	0.00575***	0.00593***	0.00553**
	(0.00148)	(0.00153)	(0.00153)	(0.00199)	(0.00219)
Unemployment 2000	-0.0260***	-0.0252***	-0.0252***	-0.0157***	-0.0364***
	(0.00465)	(0.00453)	(0.00453)	(0.00598)	(0.00950)
Infant mortality 2000	0.000211	0.000294	0.000294	5.97e-05	0.000573
	(0.000620)	(0.000671)	(0.000671)	(0.000845)	(0.00108)
Income 1910 (ln)	-0.203***	-0.211***	-0.211***	-0.0182	-0.440**
	(0.0605)	(0.0666)	(0.0666)	(0.0421)	(0.190)
Literacy 1910	-0.0662	-0.313**	-0.313**	-0.0790	-0.592*
	(0.0902)	(0.136)	(0.136)	(0.120)	(0.304)
Population 1910 (ln)	-0.00724	0.0111	0.0111	0.00241	0.0215
	(0.0111)	(0.0121)	(0.0121)	(0.0176)	(0.0205)
Manufacturing 1910	-0.0950*	-0.00892	-0.00892	-0.262***	0.292*
	(0.0509)	(0.0566)	(0.0566)	(0.0829)	(0.169)
Black population 1910	0.129	0.0901	0.0901	0.0812	0.101
	(0.0790)	(0.0848)	(0.0848)	(0.120)	(0.0950)
Female 1910	2.138***	1.587**	1.587**	0.947	2.350*
	(0.754)	(0.643)	(0.643)	(0.643)	(1.292)
Female part. 1910	0.271***	0.272***	0.272***	0.153	0.413**
	(0.0978)	(0.104)	(0.104)	(0.117)	(0.183)
Unemployment 1910	-0.142	-0.0923	-0.0923	-0.196	0.0312
	(0.165)	(0.177)	(0.177)	(0.248)	(0.291)
Observations	2,612	2,612	2,612	2,612	2,612
States	Yes	Yes	Yes	Yes	Yes
First stage F-stat	27.32	22.35	22.35	12.83	7.85

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix 8

The long-term impact of average migration distance on economic development, 1910

	OLS		IV	
Dep. Var.: income per capita 2010 (ln)	(1) All migrants	(2) Internal migrants	(3) All migrants	(4) Internal migrants
Distance all mig (ln) 1910	0.0185***		0.230***	
	(0.00360)		(0.0559)	
Distance int. mig (ln) 1910		0.0411***		0.563***
		(0.0100)		(0.135)
Education 2000	0.0116***	0.0115***	0.0103***	0.00986***
	(0.000763)	(0.000775)	(0.000991)	(0.00102)
Population 2000 (ln)	0.000480	-0.000660	-0.00289	-0.0104
	(0.00459)	(0.00484)	(0.00594)	(0.00678)
Manufacturing 2000	-0.121***	-0.124***	-0.123***	-0.0664
	(0.0237)	(0.0244)	(0.0372)	(0.0423)
Black population 2000	0.000268	6.54e-06	0.000334	-0.00156***
	(0.000273)	(0.000282)	(0.000464)	(0.000595)
Female 2000	0.0169***	0.0168***	0.0177***	0.0175***
	(0.00155)	(0.00153)	(0.00261)	(0.00222)
Female participation 2000	0.00181**	0.00210***	0.00146	0.00209**
	(0.000755)	(0.000765)	(0.00103)	(0.00102)
Unemployment 2000	-0.0213***	-0.0211***	-0.0184***	-0.0155***
	(0.00298)	(0.00312)	(0.00415)	(0.00477)
Infant mortality 2000	0.000141	0.000158	9.41e-05	0.000490
	(0.000425)	(0.000435)	(0.000604)	(0.000582)
Income 1910 (ln)	0.00532	0.0161	-0.0538**	-0.0484*
	(0.0138)	(0.0145)	(0.0263)	(0.0266)
Literacy 1910	0.149***	0.111***	0.205***	0.101
	(0.0392)	(0.0395)	(0.0681)	(0.0678)
Population 1910 (ln)	-0.00214	0.00288	-0.0210**	0.0215**
	(0.00590)	(0.00653)	(0.00952)	(0.00964)
Manufacturing 1910	-0.0805***	-0.0773***	-0.154***	0.00390
	(0.0255)	(0.0265)	(0.0426)	(0.0446)
Black population 1910	0.116**	0.126**	0.0901	0.106
	(0.0583)	(0.0585)	(0.0927)	(0.104)
Female 1910	-0.681***	-0.747***	-0.0815	-0.124
	(0.154)	(0.161)	(0.261)	(0.279)
Female participation 1910	0.118**	0.0988*	0.137	0.185**
	(0.0516)	(0.0537)	(0.0842)	(0.0864)
Unemployment 1910	0.0516	0.0611	-0.191	-0.0810
	(0.0857)	(0.0860)	(0.149)	(0.127)
Observations	2,588	2,501	2,588	2,501
States	Yes	Yes	Yes	Yes
R-squared	0.677	0.678	-	-
First stage F-stat	-	-	23.51	25.23

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1