Essays in Labor Economics and Structural Change

András Jagadits

TESI DOCTORAL UPF / Year 2022

THESIS SUPERVISOR
Joan Monràs
Department of Economics and Business

Universitat Pompeu Fabra Barcelona

Acknowledgement

First and foremost, I want to thank my parents for their unrelenting support (and sacrifices) during the previous twenty-eight years. Second, my academic progress owes its most to my supervisor, Joan Monràs, who always persuaded me to fully realize my potential. Without his guidance and our long discussions over the years, I would have a considerably more limited understanding of what it means to present, reason, write or simply think as an economist. Third, I have also learned a lot from the detailed feedback of my outstanding committee: Albrecht Glitz, Dávid Krisztián Nagy and Giacomo Ponzetto. Additionally, many people in the UPF and CREi research community have contributed to this thesis through countless seminars and discussions. Last but not least, I thank my fellow PhD students, some of whom I proudly call my friends, for their constant encouragement and support. They made the emotional roller-coaster of the last six years much more bearable and often even fun.

Abstract

This thesis studies questions at the intersection of labor economics and the literature of structural change. The first chapter empirically assesses the impact of emigration on structural change. Exploiting plausibly exogenous differences in pre-WWI chain migration from Hungary as a function of distance to the first pioneering emigrants, I find that low-skilled emigration led to local deindustrialization. To explain my empirical findings, I develop a theoretical model of a small open economy in the second chapter, whose key assumption is that manufacturing exhibits external economies of scale. In this setting, a shrinking labor force stunts industrialization when labor and land are sufficiently strong complements in agriculture. The last chapter analyzes to what extent and how individuals in occupations that are beneficially affected by structural transformations can transmit their gains in socio-economic status to their offspring. Using matching and fixed effects regressions, I document that the (grand)sons of machinists, an occupation particularly demanded in the United States during the Second Industrial Revolution, held occupations with higher earnings than the (grand)sons of comparable non-machinists. I identify rural-to-urban migration and secondary education as the main channels of intergenerational transmission.

Resumen

Esta tesis investiga temas en la intersección de dos literaturas: la economía laboral y el cambio estructural económico. El primer capítulo evalúa empíricamente el impacto de la emigración poblacional en el cambio estructural. Para ello, aprovecho diferencias plausiblemente exógenas en las cadenas de emigración desde Hungría anteriores a la Primera Guerra Mundial en función a la distancia a los primeros emigrantes. Los resultados de este capítulo muestran que la emigración poco calificada condujo a la desindustrialización local. Para explicar estos hallazgos empíricos, el segundo capítulo presenta un modelo teórico de una economía pequeña y abierta cuyo supuesto clave es que la manufactura exhibe economías de escala externas. Usando este modelo, muestro que cuando la mano de obra y la tierra son suficientemente complementarios en el sector agrícola, una reducción en la mano de obra puede frenar el proceso de industrialización de una economía. El último capítulo analiza cómo y en qué medida pueden los individuos en ocupaciones que se ven beneficiadas por las transformaciones estructurales transmitir sus logros socioeconómicos a su descendencia. Usando técnicas de matching y regresiones de efectos fijos, documento que los hijos y nietos de maquinistas (una ocupación particularmente demandada en los Estados Unidos durante la Segunda Revolución Industrial) ocuparon puestos con ingresos más altos que los hijos y nietos de no-maquinistas con características observables comparables. Los resultados identifican a la migración del campo a la ciudad y a la educación secundaria como los principales canales de esta transmisión intergeneracional.

Preface

The first two chapters of this thesis explore, empirically and theoretically, how emigration shapes local structural transformation and industrialization at an early stage of development. While studies of immigration shocks in advanced economies abound, emigration is much less studied in spite of its prevalence in the developing world, its research being limited by data availability or the lack of a credible identification strategy. Therefore, the question is still open if emigration-induced labor scarcity causes firms to purchase machines, substituting for laborers, and to increase their labor productivity, or to abandon emigration-exposed regions as non-emigrating locals are anchored by agricultural land, causing an economic decline. To address this question, the first chapter studies emigration from Hungary to the United States at the turn of the 20^{th} century. My empirical design exploits plausibly exogenous differences in chain migration as a function of distance to the first pioneering emigrants. Notably, this distance is uncorrelated with numerous growth-related characteristics and industrial growth prior to mass migration. Using newly digitized panel data on factor prices, mechanization, migration and sectoral employment, I document that low-skilled emigration led to local deindustrialization. In particular, industrial employment losses are estimated to be 2.3 times larger than the direct effect that is implied by the emigration of industrial workers themselves. Most of these losses stemmed from fewer factory openings and a slower expansion of pre-existing factories in labor-intensive sectors. In addition, I find no evidence of induced mechanization in manufacturing, while initial low-skilled wage gains disappeared despite ongoing emigration in the long run.

In the second chapter, I interpret the findings in a two-sector, small open economy model. In this model, the key difference between the two sectors is that manufacturing exhibits external economies of scale, while agriculture produces with a constant-returns-to-scale technology. If labor and land are poor substitutes in agricultural production, labor demand is relatively inelastic in agriculture, so it is mainly manufacturing that suffers from employment losses due to emigration. Consequently, scale economies weaken which exacerbates the initial manufacturing employment decline and exerts a negative effect on wages and, thus, on mechanization. Leveraging the structure of the model, I also quantify the strength of scale externalities which turns out to be similar to more contemporary estimates. In conclusion, this chapter highlights the role of external economies of scale and labor-land complementarity in agriculture in explaining the effect of migration on local development and suggests that, next to brain drain, the loss of relatively lower-skilled workers may also have severe repercussions on local structural transformation.

Likewise addressing the intersection of labor economics and economic growth in the third chapter (a joint work with Laurenz Bärtsch), we try to understand to what extent, and how individuals in occupations that are beneficially affected by structural transformations can transmit their gains in socio-economic status to their (grand)children. While this question is very timely considering, for instance with the current wave of automation, the analysis of recent shocks is inhibited by the time horizon required for intergenerational research. Therefore, we analyze the case of machinists whose occupation experienced a relative labor demand spike during the Second Industrial Revolution (1870-1914), resulting in higher income and job stability. To do so, we complement data from the US full count census with newly digitized data on the county-level supply of secondary education and occupation-state level earnings. Using matching and fixed effects regressions, we document that the (grand)sons of men who were machinists in 1870 held occupations with significantly higher earnings than the (grand)sons of comparable non-machinists. The higher earnings of machinists' sons mainly stemmed from parental investment in their education, but this effect is absent for those sons who were already too old to attend high school when the income of machinists started to rise. Additionally, the sons of initially rural machinists benefited from rural-to-urban migration. Our results are robust to controlling for family-fixed effects (comparing machinists to their non-machinist brothers), pre-1870 spatial sorting, and a rich set of next-door neighbor and grandparental characteristics. Our findings suggest that the effects of current transformations in the labor market might be passed on to later generations, but to a lesser extent owing to a considerably more expanded public education system compared to the turn of the 20$^{\text{th}}$ century.

Contents

List of Figures XIII

List of Tables XVII

1. EMIGRATION AND LOCAL STRUCTURAL CHANGE EMPIRICAL EVIDENCE FROM (AUSTRIA)-HUNGARY IN THE AGE OF MASS MIGRATION 1

1.1. Introduction 1
1.2. Historical background 4
1.3. Data 11
 1.3.1. Data on emigration 11
 1.3.2. Data on industrial employment 12
 1.3.3. Data on wages 12
 1.3.4. Data on mechanization and capital stock 13
 1.3.5. Other data sources 14
 1.3.6. Sample construction 14
1.4. Empirical strategy 16
 1.4.1. Regression specification 16
 1.4.2. Instrumental variable 17
 1.4.3. Correlation with observable determinants of growth 20
 1.4.4. Pre-trend analysis 20
1.5. Main results 22
 1.5.1. First stage 22
 1.5.2. County-level estimation 22
 1.5.3. Factory employment - municipality-industry subgroups 28
1.6. Additional empirical results 29
 1.6.1. Output mix 29
 1.6.2. Low-skilled wage growth 32
 1.6.3. Employment and mechanization in manufacturing 34
 1.6.4. The capital stock of public limited companies 35
 1.6.5. The evolution of regional prices 36

1.6.6. Heterogeneous effect by the initial level of industrialization in municipalities 37
1.7. Conclusion 38
A. Appendix - Chapter 1 38
A.1. Data appendix 38
A.2. Additional anecdotal evidence and empirical analysis . . . 45
A.3. Additional figures 57
A.4. Additional descriptive statistics 66
A.5. Additional regression tables 69

2. EMIGRATION AND LOCAL STRUCTURAL CHANGE THEORY **77**
2.1. Introduction 77
2.2. Main assumptions 80
2.2.1. Discussion of the main assumptions 80
2.2.2. Primitives of the economy 82
2.3. Equilibrium 83
2.4. Emigration in the model 84
2.5. Extensions 86
2.6. Quantifying the scale externalities 89
2.7. Conclusion 91
B. Appendix - Chapter 2 92
B.1. A literature survey of agglomeration economies in historical contexts 92
B.2. Equilibrium stability and uniqueness 93
B.3. Equilibrium conditions following Jones (1965) 95
B.4. Detailed derivation of the equilibrium conditions 97
B.5. Proof of Propositions 1 and 2 98
B.6. Proof of Proposition 3 99
B.7. Proof of Proposition 5 99

3. THE MECHANICS OF GOOD FORTUNE ON INTERGENERATIONAL MOBILITY DURING THE SECOND INDUSTRIAL REVOLUTION **101**
3.1. Introduction 101
3.2. Historical background 106
3.3. Data 108
3.3.1. Linking historical censuses 109
3.3.2. Occupational earnings scores for the late nineteenth century 110
3.3.3. Summary statistics 111
3.4. Empirical strategy 112

3.4.1. Propensity score matching . . . 112
3.4.2. Fixed effects regression . . . 115
3.5. Main results . . . 116
3.5.1. Long-term effects and intergenerational transmission . . . 116
3.5.2. Mechanisms behind the intergenerational transmission . . 124
3.5.3. Fathers in other demanded occupations . . . 131
3.6. Robustness checks . . . 132
3.6.1. Robustness checks using matching . . . 133
3.6.2. Robustness checks using regressions . . . 136
3.6.3. Grandfather-fixed effects . . . 138
3.6.4. Correcting measurement error and magnitude comparison 141
3.7. Conclusion . . . 141
C. Appendix - Chapter 3 . . . 142
C.1. Data appendix . . . 142
C.2. Inverse proportional weights . . . 150
C.3. Additional empirical results . . . 150

List of Figures

1.1. Emigration to the United States (1871-1913) 7
1.2. Emigration exposure (US) at the county level (1900-1910) 9
1.3. Emigration exposure (US) 19
A1. Emigration to the United States by sending countries (1821-1910) 57
A2. The counties of Hungary before WWI 58
A3. Emigration to the United States (1871-1913) 59
A4. Emigration and the Hungarian economy (1871-1913) 60
A5. Emigration and agricultural production (1900-1912) 60
A6. Yearly emigration and wheat production at the county level (1901-1910) 61
A7. The expansion of Hungarian railway lines 62
A8. The spatial distribution of German-ethnicity citizens in 1900 . . . 62
A9. First stage - county level 63
A10. Reduced form - county-level factory employment change 64
A11. Reduced form - municipality-industry subgroup level factory employment growth 65

2.1. The effect of emigration 86
B1. The pre-industrial equilibrium is the only equilibrium 94
B2. Corner equilibria are the stable equilibria 95
B3. Equilibrium stability and uniqueness 96

3.1. The evolution of occupational employment shares over time . . . 107
3.2. Histogram of occupational employment changes 107
3.3. The histogram of some continuous characteristics after matching . 114

List of Tables

1.1. Observable characteristics - emigrants and total population 10
1.2. Summary statistics - municipality-industry subgroup sample (1900-1910) 16
1.3. Pre-trends and the period of interest - main outcomes 21
1.4. First stage - county-level analysis 23
1.5. Population and worker losses (1900-1910) 24
1.6. Decomposition of worker losses (2SLS; 1900-1910) 25
1.7. Disproportionately large industrial worker losses (1900-1910) . . 26
1.8. Factory employment losses (1900-1910) 27
1.9. Municipality-industry subgroup level employment growth (1900-1910) 29
1.10. Decomposition of factory employment losses by labor intensity (1900-1910) 31
1.11. Agricultural and manufacturing wage growth (1898-1912) 33
1.12. Engine power capacity and employment growth (1901-1912) . . . 35
1.13. Decomposition of factory employment losses by the local level of industrialization (1900-1910) 37
A1. Capital and labor-intensive sectors - 1913 (Great Hungarian Compass) 43
A2. Foreign trade of Hungary - value share of main categories 47
A3. Population growth decomposition (1900-1910) 50
A4. Net internal out-migration decomposition (1900-1910) 50
A5. Robustness checks - county-level analysis 52
A6. Change in the number of factory workers - two instrumental variables 53
A7. Municipality-industry subgroup level sample - nearest neighbor matching 53
A8. Change in the supply and price of financial capital 54
A9. High- and low-skilled employment growth (1901-1912) 56
A10. Summary statistics - county sample 66
A11. Largest industries in the municipality-industry subgroup sample . 67
A12. Largest expansions and contractions - full sample and municipality-industry subgroup panel 67

A13. Dispersion of retail prices and inflation rates - regional markets (1900-1910) 68
A14. Placebo regressions 69
A15. First stage - municipality-industry subgroup level analysis 70
A16. Disproportionately large industrial worker losses (1900-1910) . . 70
A17. Factory employment losses and firm dynamics 71
A18. Municipality-industry subgroup level employment growth - sectoral split 72
A19. Decomposition of factory employment losses by tradability (1900-1910) 72
A20. Agricultural wage growth (1898-1912) - subcounty-level analysis . 73
A21. Industrial inspector reports - basics 73
A22. Engine power capacity and employment log-levels (1901) 74
A23. Decomposition of capital stock changes at public limited companies (1900-1912) 74
A24. Factory employment losses - additional IV for the level of industrial employment in 1891 75

2.1. Scale elasticity estimates 90

3.1. Occupational earnings (1850-1892; in 1890 dollars) 109
3.2. Summary statistics of fathers (G1 in 1870) 112
3.3. Main outcomes - fathers (G1; 1870-1900) 117
3.4. Measures of economic status (medium- and long-run) - fathers (G1) 119
3.5. Main outcomes - sons (G2; 1900) 121
3.6. Measures of economic status - sons (G2; 1900) 122
3.7. Main outcomes - grandsons (G3; 1940) 123
3.8. Measures of income - grandsons (G3; 1940) 124
3.9. Heterogeneity by the level of private tuition fee - sons (G2; 1900) 126
3.10. Heterogeneity by the supply of public schooling - sons (G2; 1900) 128
3.11. Information channel - sons (G2; 1900) 130
3.12. The urban-rural gap in the earnings effect (G2; 1900) 132
3.13. Sons of fathers in other occupations (G2; 1900) 133
3.14. Within-family estimation - fathers (G1; 1870-1900) 140
C1. Top control occupations 151
C2. Migration destination decomposition - fathers (G1; 1900) 151
C3. Measures of economic status - fathers (G1; 1880) 152
C4. Measures of occupational income and mobility - occupation switcher fathers (G1; 1870-1900) 153
C5. Measures of education and wealth - sons (G2; 1940) 154
C6. Main outcomes - sons (G2; 1900) 155

C7. Measures of economic status - sons (G2; 1900) 156
C8. Robustness checks - sons (G2; 1900) 156
C9. Main outcomes - fathers (G1; 1870-1900) 156
C10. Robustness checks with regressions - sons (G2; 1900) 157
C11. Ability bias - fathers (G1 in 1870) 157
C12. Robustness checks by the age of fathers and sons - sons (G2 in 1900) 158

Chapter 1

EMIGRATION AND LOCAL STRUCTURAL CHANGE

EMPIRICAL EVIDENCE FROM (AUSTRIA)-HUNGARY IN THE AGE OF MASS MIGRATION

1.1 Introduction

Structural change, the reallocation of labor from agriculture towards modern sectors, is a hallmark of economic growth. While developing countries are undergoing this process, many of them are losing population due to emigration to more affluent countries (Dao et al., 2018; Clemens, 2020). Since the majority of these developing economies are still at an early phase of industrialization, the question of whether emigration has an impact on the speed of structural change arises. Yet previous studies do not offer a definitive answer.

On the one hand, the migration literature typically finds that (low-skilled) immigrant-induced labor supply shocks primarily lead to the substitution of capital with labor *within* sectors rather than any direct effect on structural change (see Lewis, 2013 for a literature review). This would imply a tenuous link between emigration and structural change. On the other hand, prior literature suggests that a labor surplus in agriculture might be a prerequisite for the development of modern sectors (e.g., Lewis, 1954; Gollin et al., 2002, 2007; Bustos et al., 2016; Leukhina and Turnovsky, 2016). For instance, when labor-saving agricultural technologies are adopted, agricultural workers move to manufacturing and this helps the manufacturing sector develop. This evidence would suggest that the emigration of agricultural workers can slow down structural transformation. Labor shortages in agriculture may drive manufacturing workers towards agriculture, curbing the development of the sector.

To investigate how emigration shapes local structural transformation, I study mass emigration from Hungary to the United States at the turn of the 20th century. Over the years 1899-1913, approximately 1.2 million citizens, mostly unskilled agricultural workers, left the country and merely one-quarter of them returned, resulting in a 4-5% population loss. Anecdotal evidence suggests that emigration-induced labor shortage emerged as the main hurdle to industrialization in these years (MGYOSZ, 1907), when Hungary was still at the beginning of a transition to become an industrialized economy.[1]

Using newly digitized panel data sets on manufacturing employment and engine power capacity, firm balance sheets, low-skilled wages and migration (international as well as internal), I exploit the fact that regions close to the county from where the first pioneering emigrants left Hungary sent persistently more emigrants to the US. Various sources suggest that this phenomenon was a result of local information diffusion and social network-driven chain migration from regions in geographic proximity to earlier emigrants in current southern Poland rather than the consequence of negative economic shocks.[2] Hence, geographic distance from the first Hungarian pioneers is used as an instrumental variable for the possibly endogenous local emigration rate. I present reduced form evidence on the absence of pre-trends and demonstrate that the instrumental variable is not correlated with predetermined potential drivers of economic growth. These empirical exercises support the identification assumption that, in the absence of mass migration, outcomes would have evolved similarly across high- and low-emigration counties.

Leveraging this empirical strategy, I demonstrate that emigration led to local deindustrialization. First, I show that county population significantly decreased as a consequence of emigration to the US, indicating a negative labor supply shock on spatially segmented local labor markets. Second, industrial employment losses (primarily manufacturing and mining) stand out despite the agricultural background of most emigrants. More precisely, I observe a relative decline of 2.3 industrial workers in the local population for every industrial worker who emigrated. Finally, the majority of these industrial employment losses stemmed from modern factories (plants with more than twenty employees) rather than smaller workshops or individual craftsmen.

[1] Crafts and manufacturing provided almost one-fifth of the national income on the eve of WWI, but approx. two-thirds of the labor force was still employed in agriculture (Schulze, 2007a; Klein et al., 2017). The GDP per capita of Hungary around 1910 was close to modern-day Bangladesh, Cambodia, Ghana, Kenya or Tanzania (Bolt and van Zanden, 2020).

[2] I further corroborate the importance of this pull factor by showing that yearly agricultural production and emigration were uncorrelated in the studied time period. This indicates that, for instance, a bad harvest did not act as a push factor for emigration.

To shed light on the mechanism behind the relative industrial downturn, I show that the negative labor supply shock did not induce capital-intensive technology adoption: the negative impact on engine power capacity growth was even larger than on employment growth in manufacturing. Additionally, I find a negative effect on the the capital stock of public limited companies. This effect is primarily driven by the labor-intensive branches of industry. In line with this finding, a sectoral employment decomposition reveals that labor-intensive manufacturing and mining experienced severe losses, while the effect on capital-intensive manufacturing was less marked but also negative. I do not find a similar significant difference between more and less tradable sectors. In addition, I present evidence that the long-term effect on agricultural and industrial low-skilled wages was negative.

In conclusion, emigration exerted a negative impact on local structural transformation: industrial employment shrank mostly in labor-intensive sectors, without an ensuing capital deepening and more than the change predicted by the sheer loss of emigrated industrial workers. While this chapter exclusively documents these pieces of empirical evidence, the next chapter offers a theoretical model to rationalize the findings and discuss their external validity.

Related literature This work primarily speaks to the literature that studies the effect of emigration on growth.[3] Hornbeck and Naidu (2014) analyze the Great Mississippi Flood of 1927 and show that black out-migration caused modernization and higher capital intensity in the agriculture of affected areas. Clemens et al. (2018) find a similar effect after the exclusion of Mexican laborers from the US.[4] While these two papers examine the impact of out-migration on agriculture, this chapter focuses on the effect on industry. Another recent working paper which does so is Andersson et al. (2020). They find evidence for the beneficial impact of emigration to the US in late-nineteenth-century Sweden. Places more exposed to emigration benefited from more patents, modern technology adoption and faster structural change. However, several authors observe an opposite effect in the US, where the Quota Acts drastically reduced immigration in the 1920s. Consequently, labor markets relying on abundant cheap labor from Austria-Hungary, Italy or Russia experienced a negative labor supply shock. In these places, diminished

[3]The broader literature shows that emigration may result in brain gain (Mayr and Peri, 2009; Dustmann et al., 2011). If emigration opportunities are better for the higher-skilled, this might provide incentives to education at the origin (Beine et al., 2008; Shrestha, 2017). On the flip side, the loss of high-skilled individuals may stunt growth (Docquier and Rapoport, 2012). Since most Hungarian emigrants were low-skilled workers with an agricultural background, human capital losses are unlikely to drive my findings (see Section 1.6.2). Additionally, repatriated savings and remittances can improve well-being in sending locations (Yang, 2011; Gibson and McKenzie, 2014; Abramitzky et al., 2019b).

[4]Lafortune et al. (2015) find significant output mix changes in agriculture as a result of immigration in the early-twentieth-century United States.

low-skilled immigrant inflows led to i) fewer patents (Arkolakis et al., 2019;[5] Doran and Yoon, 2020), ii) smaller average establishment size and less intensive electricity adoption (Xie, 2017; Tabellini, 2020), and iii) reduced labor productivity in manufacturing (Ager and Hansen, 2017; Tabellini, 2020).[6] My empirical results are in line with these findings. My contribution to this literature, combined with Chapter 2, is showing that emigration can have a negative impact on growth through industrial scale externalities for countries in an early phase of industrialization.

Additionally, this chapter speaks to at least two strands of the economic history literature. First, it is related to the growing literature that studies economic phenomena related to the Age of Mass Migration. While the overwhelming majority of papers focus on the United States (e.g., Abramitzky et al., 2012, 2013, 2019b; Lafortune et al., 2019; Sequeira et al., 2020; Tabellini, 2020), the literature addressing the effect in sending countries is very limited. Karadja and Prawitz (2019) find that emigration from Sweden caused increased local demand for political change as measured by labor movement membership, strike participation, and voting in the 19$^{\text{th}}$ century. Meanwhile, Fernández-Sánchez (2020) shows that emigration from Spain led to more intensive school construction and diffusion of beliefs about the value of education and effort. Second, this work is connected to recent studies of the early phases of industrialization. This chapter reinforces the view that plentiful local labor was conducive to growth during the Second Industrial Revolution (e.g., Kim, 2006, 2007; Martínez-Galarraga, 2012; Doran and Yoon, 2020). In doing so, it also shows that not only the highest skilled locals mattered for development in certain early stages of industrialization (Kelly et al., 2014; Squicciarini and Voigtländer, 2015).

The chapter is structured as follows. Section 1.2 explains the historical background of mass migration from Hungary. Section 1.3 describes the data sources and estimation sample construction with summary statistics. Next, Section 1.4 discusses the empirical strategy. Section 1.5 contains the main empirical results and Section 1.6 investigates potential mechanisms empirically. Section 1.7 concludes.

1.2 Historical background

This section describes the historical background of (Hungarian) mass emigration to the United States and provides some summary statistics on the characteristics of emigrants with special attention to their occupation.

[5] Arkolakis et al. (2019) focus on the period between 1880 and 1920 instead of the post-Quota Acts period.

[6] Xie (2017) finds evidence of an increasing horse power per wage earner ratio as a result of reduced low-skilled immigration as well. Ager and Hansen (2017) hypothesize the presence of agglomeration economies.

Emigration to the United States Mass migration to the US started in the second quarter of the 19th century in north-western Europe. Prior literature has already identified the major drivers of mass migration in this time period: network-driven chain migration, demographic pressures and differences in real wages and in employment ratios between sending countries and the US (e.g., Hatton and Williamson, 1994; Hatton, 1995; Spitzer and Zimran, 2019). As wages in north-western Europe started to catch up with those in the US and demographic pressures ebbed, the migratory fever shifted to more southern and eastern regions where networks facilitating emigration spread rapidly. The nationality of immigrants changed over time accordingly (Figure A1). Between 1821 and 1890, 29.2% of immigrants were German, 40.4% British and only 2.8% from Austria-Hungary. These figures are in stark contrast to the later period (1901-1910): 23.8% of immigrants were Austro-Hungarian, 22.7% Italian and only 3.8% German or 9.6% British. In the United States, the government gradually increased the barriers to immigration over the period of Mass Migration, but these measures were mostly marginal.[7] Similarly modest attempts to regulate emigration were also made by the Hungarian government.[8] Thus, the entire studied time period can be treated as an age of practically unrestricted migration.

The first Hungarian pioneering emigrants left from the northern county of Sáros (see Figure A2) in 1862. However, the flow took decades to gain importance and was almost exclusively concentrated in Sáros and the neighboring Szepes county until the 1880s. According to Gibson and Lennon (1999), there were 3,737 Hungarian-born people[9] in the US in 1870 and only 11,526 one decade later. This

[7] In 1885, the Alien Contract Labor Law outlawed the importation and immigration of foreigners under contract or agreement to perform labor in the US. The Immigration Act (1891) prohibited the immigration of those who had a contagious disease, were polygamists or likely to become public charges. In 1903, the Anarchist Exclusion Act banned anarchists, beggars and importers of prostitutes from immigrating. The Immigration Bill of 1907 increased the immigrant head tax to 4 dollars (originally set as 50 cents in 1882). The US started to require immigrants over the age of 16 to demonstrate basic reading ability in any language in 1917. Sources: HRCSO (1918) and https://www.pewresearch.org/fact-tank/2015/09/30/how-u-s-immigration-laws-and-rules-have-changed-through-history/ - accessed on 18/12/2019.

[8] Emigration agents were required to have an official permission which only Hungarian citizens could get in exchange for a lump sum deposit (1881). Thus, the Hungarian government could control the number of *legal* agents. In practice, this regulation was easily circumvented and the government still accepted the citizens' freedom of movement more than two decades later. Leading politicians acknowledged that the vast gap in earnings between the US and Hungary could not be closed within a short time span. Therefore, any emigration ban would have been futile. Instead, the main aim of Hungarian migration policy was to deter citizens from dangerous destinations necessitating the intervention of the Hungarian state and to make sure that the majority of emigrants, especially the patriotic ones with savings, returned. Sources: Pálvölgyi (2010); Poznan (2017); Pálvölgyi (2018).

[9] I cannot isolate those Hungarian citizens who were born in Croatia-Slavonia either in the US

number increased by roughly 350,000 between 1900 and 1910.[10] This pattern can also be observed in combined European port data in Figure 1.1. In the decades of pioneering, emigrants were more likely to leave for (recurrent) temporary work in the US rather than with the goal of permanent emigration (followed by the rest of their family), which became the norm in the fifteen years before World War I.[11] As opposed to earlier emigrants from north-western Europe, south-eastern Europeans were employed in the booming US manufacturing and mining sectors as the expansion of the American frontier, which previously provided easy access to land, slowly ended by these years.

Figure 1.1 shows that Hungarian migration to the US before the late 1890s is dwarfed by the numbers in the first decade of the 20th century and especially by the period between 1905 and 1907. Contemporaries and historians have argued that the upswing in migration is related to the US business cycle (Hegedüs, 1911; Jerome, 1926; Puskás, 1983). The US economy experienced a serious crisis between 1893 and 1897 which reduced labor demand. As the economy recovered and a boom started in 1897, the number of immigrants rapidly increased. Some years later, «*due to the unionization of local workers and their need for higher wages in 1905-1907, there was great demand for cheap labor … thus, immigration to the United States reached its maximum*» (Hungarian Royal Central Statistical Office (HRCSO), 1918, p. 15*).[12] Apart from the US business cycle, the intensifying competition between shipping companies drove down the cost of emigration in these years. The option to emigrate from the Hungarian port of Fiume (currently Rijeka) directly to New York became available in 1904 as well. The opening of this port reduced transportation costs for Hungarians, who no longer had to emigrate through faraway German ports.[13]

In contrast to the documented effect of the US business cycle, deteriorating economic conditions in Hungary cannot explain the rapid rise in emigration. As

census or in European port data. There were also some Hungarian-born people in the US who emigrated after the Revolution of 1848-49.

[10] HRCSO (1918) argues that this number must be a vast underestimate which originated from the inaccurate processing of place of birth for Croatian- and Slovak-ethnicity Hungarian citizens.

[11] Repeated migration results in the overestimation of the net number of emigrants before 1900 because the same person travelled again and because of simple return migration. See the first volume of the Hungarian census in 1891 (p. 70-73.) on the temporary nature of emigration in the first decades. Fewer than 0.6% of Hungarian citizens were outside Austria-Hungary according to this census.

[12] Jerome (1926, p. 121) writes that «*business conditions are in fact a dominating determinant of cyclical fluctuations in immigration. The influence of a major cyclical change in industrial conditions is usually apparent in immigration within less than a half year*». Wyman (1993) notes unusually high return migration after the Panics of 1893 and 1907.

[13] The main ports used by emigrants were Bremen, Hamburg, Antwerp and eventually Fiume/Rijeka. Out of the 1.2 million emigrants, 0.52 million were transported via Bremen, 0.21 via Hamburg and 0.24 via Fiume/Rijeka between 1900 and 1910.

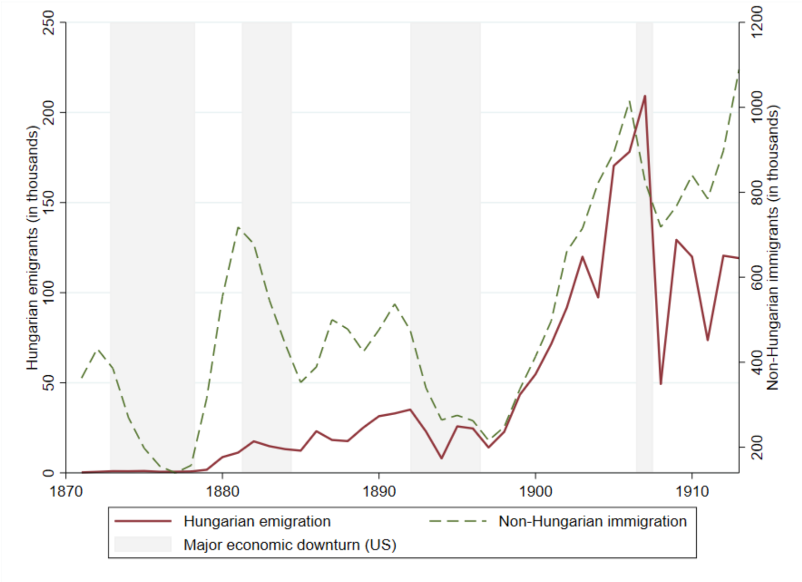


Figure 1.1: Emigration to the United States (1871-1913)

Note: the source of the number of Hungarian emigrants (left axis) is combined European port data published in HRCSO (1918, p. 150). These data were published in calendar years and also include the citizens of the autonomous Croatia-Slavonia. Data on the total number of immigrants to the US are from the Yearbooks of Immigration Statistics. They were recorded in fiscal years ending on June 30. To convert fiscal years into calendar years, I assign the mean of the fiscal year ending and starting in a given calendar year to every calendar year. The values on the right axis are constructed as the difference between the total number of immigrants to the US and of Hungarian emigrants to the US (US data only contain combined, Austro-Hungarian immigration in some years within the period of interest). A recession period qualifies as major downturn if business activity declined by at least 25% as calculated by Zarnowitz (1996). The Panics of 1893 and 1896 are treated as one. An alternative plot of Hungarian versus total immigration to the US, avoiding the fiscal year to calendar year conversion, can be found in the Appendix (Figure A3).

contemporaries wrote, «*if we seek to find the economic roots of emigration, economic conditions around 1905 contradict the abnormal increase in emigration. The harvest was decent in 1905 and excellent in 1906 ... the financial system was improving as well*» (HRCSO, 1918, p. 15*).[14] Recent publications on Hungarian mass migration agree with historical sources and conclude that «*emigration was primarily influenced by pull factors...we are not aware of any major change in domestic circumstances and the literature supports us in this claim*» (Kulcsár and

[14]Mailáth (1900) notes the strong (weak) connection between US (Hungarian) labor demand and Hungarian emigration, too.

Kulcsár, 2019, p. 93).

I present additional evidence for these claims in Figure A4, which shows the output index of Hungarian manufacturing together with the time series of emigration. The data show that Hungary experienced an unprecedented manufacturing expansion in the 10-15 years prior to WWI. I jointly plot the share of ploughing land devastated by natural forces (frost, drought, etc.) which suggests that there were no large, negative agricultural shocks driving emigration in the period of interest. In particular, the share of ploughing land destroyed never exceeded 5% significantly. As variation in agricultural output can also come from the intensive margin of production, I show that the output index of five major grains (Eddie, 1968) was not correlated with the level of emigration in Figure A5. Finally, in Figure A6, I establish the lack of any significant correlation between the yearly production of the most important grain (wheat) and emigration at the county level. Bodovics (2019), who focuses on the region of pioneers, also concludes that chain migration was already a dominant factor in the 1880s and that the size of yearly pioneering emigrant cohorts was not related to the quality of the harvest.

Other international destinations Around 80% of all Hungarian emigrants who left Austria-Hungary migrated to the US, making it the single most prominent destination, but there were two other foreign destinations which are worth mentioning: Romania and Germany. People living in Transylvania (currently north-western Romania) chose the Romanian direction, mostly in search of seasonal agricultural work in the relatively less densely populated country. The start of (temporary) emigration in this direction dates back to centuries before the Age of Mass Migration. Emigration to Germany, on the other hand, was fueled by the labor demand of the rapidly industrializing country. Migration to the also comparatively rapidly developing Austria, which most likely surpassed the sum of out-migration to Germany and Romania in magnitude, was quasi-internal migration in the Austro-Hungarian era. Thus, the best available data source on migration to Austria are census data collected by Austrian officials. Clear patterns observed by HRCSO (1918) are that the importance of this direction diminished from the west to the east and as emigration to the US spread out.

The extent and composition of Hungarian Mass Migration to the United States To analyze the extent of emigration-induced population losses, emigration exposure is defined as the number of emigrants to the US from every Hungarian county between 1900 and 1910, adjusted for return migration in the same period, and divided by county population in 1900. It is also multiplied by 100 so that it can be interpreted in percentage terms. The distribution of emigration exposure can be seen in Figure 1.2. While around 40% of counties did not experience mass

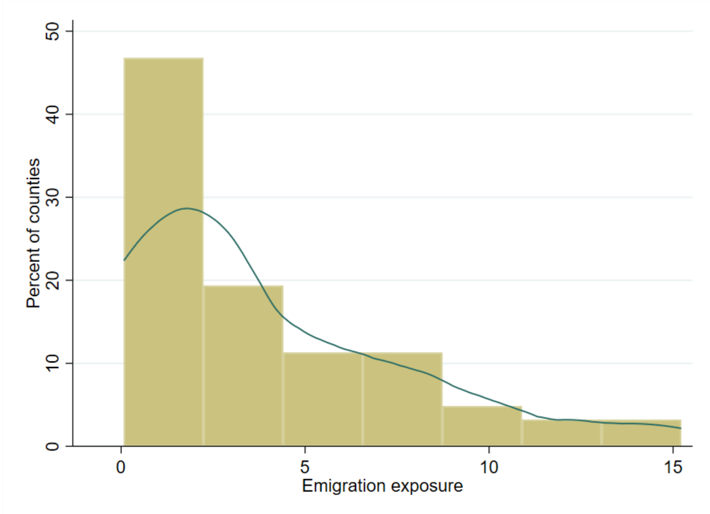


Figure 1.2: Emigration exposure (US) at the county level (1900-1910)

Note: emigration exposure is defined in the main text.

emigration, others lost more than 10% of their initial population. Row 1 of Table A10 shows that the mean value of emigration exposure is 3.5%.

HRCSO (1918) reports that the occupational composition of yearly emigrant flows was rather stable over time. The majority of emigrants (54.8%) were day laborers or servants working in agriculture (Table 1.1). Meanwhile, the share of these occupational categories among all Hungarian workers was merely 23% in the 1910 census.[15] Freelancer day laborers (not associated with either the industrial sector or agriculture) were also overly represented in the pool of emigrants, relative to their share in the labor force. Owner-occupier farmers (i.e. those who owned a plot) emigrated below their occupational share, which is consistent with Abramitzky et al. (2013) who show that wealth was a negative predictor of emigration in Norway during the Age of Mass Migration. Likewise, individual craftsmen (for instance, self-employed blacksmiths, shoemakers, tailors, etc.) and industrial laborers were underrepresented among emigrants compared to their share in the labor force. Moreover, the number of workers in high-skilled, liberal professions (e.g., lawyers, physicians, scientists, teachers, etc.) was barely affected by the emigration of these individuals. In conclusion, Hungarian Mass Migration predominantly involved low-skilled laborers.[16]

[15]Officials collecting data on emigration and statisticians at the Central Statistical Office used slightly different occupational categories which were only harmonized for some tables of HRCSO (1918). Therefore, I can exclusively rely on their published comparisons with the 1910 census.

[16]The share of non-Hungarians and of women is also noteworthy. HRCSO (1918) observes that mixed-ethnicity municipalities lost non-Hungarian- as well as Hungarian-ethnicity population.

Table 1.1: Observable characteristics - emigrants and total population

Category	Share of emigrants (1905-07; %)	Share of workers or population (1910 census; %)
Occupation		
Day laborer (agriculture)	54.8	23.0
Owner-occupier farmer	10.6	38.1
Individual craftsman	2.3	5.6
Laborer (industry)	11.8	14.7
Day laborer (no agr./ ind.)	11.1	2.6
Liberal professions	0.5	3.9
Other	8.9	12.1
Gender		
Female	29.5	50.4
Male	70.5	49.6
Ethnicity (mother tongue)		
Hungarian	33.9	54.4
Other	66.1	45.6
Age		
Below 20 years old	24.1	45.4
20-29 y.o.	35.4	15.6
30-39 y.o.	25.3	12.0
40-49 y.o.	12.6	10.3
Above 50 y.o.	2.6	16.7

Note: partial reproduction of tables in HRCSO (1918, p. 18*, 26*, 29* and 34*) which does not include some minor occupational categories. These data were published for 1905-07 and include all emigrant destinations outside Austria-Hungary. The share of workers from the 1910 census is used for the occupational comparison in the top panel, while other comparisons below use the share in total population.

However, there are some caveats about the interpretation of these occupational facts. First, occupations were only recorded for heads of household or individual emigrants. Fortunately, this is not a major problem since most emigrants left individually.[17] Second, another issue is that basically nobody was recorded as dependent or without any occupation. This may be due to emigrants being categorized as agricultural workers even if they only briefly helped their families during the harvest season. Data from Ellis Island support this claim: 25.3% of all Hungarian-ethnicity immigrants were recorded as having no occupation, including

Therefore, the emigration of non-Hungarians was most likely not the result of local tensions rooted in ethnicity. The large share of non-Hungarians had rather to do with the Slovak-dominated region of pioneers and the early migration of Germans, discussed in Section 1.4.2. Importantly, the proposed instrumental variable is uncorrelated with the share of non-Hungarian ethnicity locals in my preferred specification (see Section 1.4.2). The low share of women among emigrants was partly compensated by their even lower share in return migration: while 29.5% of emigrants were women between 1905-07, only 15.4% of return migrants were females. Consequently, net emigration was substantially more balanced on gender than gross emigration statistics may suggest.

[17] For instance, approx. 397,000 emigrants left Hungary for the US between 1905-07. The occupation of 85-86% of these emigrants is known. The share of non-individual emigrants fluctuated around 10% in the period of interest.

housewives and children, between 1900 and 1913 (HRCSO, 1918, p. 71). Finally, occupational statistics on return migrants cannot be used as the data sometimes refer to a person's occupation before emigration and sometimes to their occupation in the US.

1.3 Data

In what follows, I describe the data sources and discuss their main strengths and limitations. In addition, I explain the construction of the final estimation samples and highlight what we can learn from summary statistics about the development of Hungary ca. 1900.

1.3.1 Data on emigration

The essential data source on emigration is HRCSO (1918). This publication contains the number of emigrants and return migrants (both to the US and other directions) for all counties of Hungary between 1899 and 1913. I use these data on the US direction in the main empirical analysis. HRCSO (1918) also provides information on the age, gender, ethnic and occupational distribution of emigrants.

Importantly, data on the occupational distribution of emigrants were published exclusively for 1905-07 and 1911-13. I use the former to impute the number of emigrants who were industrial workers between 1900 and 1910. I assume that their share was exactly the same as at the peak in 1905-07, when roughly half of the emigrants left in the period of interest. This is not a strong assumption since their share in 1905-07 is almost the same as in 1911-13, though slightly decreasing over time. Unfortunately, there is no reliable information on the occupation of return migrants. Thus, in the baseline analysis, I assume that the share of return migrants who took up industrial jobs after their return to Hungary equals the share of industrial workers in all emigrants in a given county.[18]

[18]This is a realistic assumption since more than 40% of return migration happened in 1907-1908, following a US economic downturn, in the period of interest. This potentially indicates that migrants could not save enough money to buy an own plot or pay back the mortgage on their plot. Therefore, they continued working as an agricultural or industrial employee. Poznan (2017) also reports that many return migrants (around 30% in a non-representative survey) did not have any or at most modest savings. Those who accumulated enough wealth, on the other hand, could start an industrial business. Valuable experience in the US industry could lead even talented, former agricultural workers to industrial sectors. The Hungarian governmental campaign to attract more return migrants was also specifically targeted at skilled industrial workers (Poznan, 2017). Abramitzky et al. (2019b) document that Norwegian return migrants bought own plots or exploited their urban experience in industry (e.g., carpentry).

In addition to Hungarian administrative data, two other data sets on migration exist: European port data and data collected on Ellis Island. These three emigration databases each have different shortcomings. First, European port data contain the citizenship of emigrants but no information on possible return migration, occupation or exact place of living in the origin country. Second, the occupation of all immigrants was recorded in the Ellis Island data set, but US officials switched from issuing data for Austria and Hungary separately to merging them as Austria-Hungary between 1899 and 1907. Moreover, US authorities maintained a vague occupational label (*worker*) which covered 26.6% of all Hungarian-ethnicity immigrants between 1900 and 1913. Finally, as discussed previously, I primarily use Hungarian administrative data on emigration, which were collected separately from the census. They provide more accurate estimates of the number of emigrants and information at the county level. However, these data suffer from the drawbacks previously described, namely that the occupation of companions of heads of family is unknown and that county-level occupational shares are reported exclusively for certain years.

1.3.2 Data on industrial employment

One of the main outcome variables, total employment in industrial establishments with more than twenty employees (which are referred to as factories), comes from a special industrial survey in the 1900 and 1910 censuses. The data set reports the number of factories and the total employment in them in every municipality-industry subgroup pair.[19] Besides the dominant manufacturing, mines, firms in construction, hospitality (mainly hotels and spas) and utilities (power plants, sanitation firms and waterworks) are also included because they were generally considered as part of industry (see a longer discussion in Appendix A.1.1). Data on broader measures of industrial employment (e.g., industrial employment including craftsmen and employees of smaller firms) come from decennial censuses as well.

1.3.3 Data on wages

I digitized the wage time series of adult male agricultural day laborers (by season) from the yearly publication of the Royal Ministry of Agriculture (*Mezőgazdasági munkabérek Magyarországon az 1XXX. évben*). Wages were published

[19] As municipality names were «hungarianized» in the period of interest, I manually checked potential name changes for all observations which could not be matched between 1900 and 1910 to see if the cause of a non-match is entry/exit or simply changing municipality name. There were no significant changes affecting county borders and expanding cities did not unite with surrounding smaller municipalities containing factories. For the latter, the only important exception is Győr where I treat factories in engulfed neighboring settlements as part of the city in the entire analysis.

at the county and subcounty level starting from the mid-1890s. Additionally, I construct a panel of average male wages in manufacturing from two sources. For 1900, I use the 6th volume of the census which reports the wage distribution in intervals for low-skilled workers (i.e. laborers paid in weekly wages) by industry group-county pairs. For 1910, I use a special survey of all factories in Hungary publishing the wage distribution of low-skilled males at the county level (*A Magyar Szent Korona országai gyáriparának üzemi és munkás-statisztikája az 1910. évről*). Average wages are calculated as employment weighted averages of the midpoints of published wage intervals in both cases. Industrial wages exclusively pertain to manufacturing - construction, hospitality and mining are not included.

1.3.4 Data on mechanization and capital stock

Yearly data on mechanization and employment by skill level can be retrieved from the reports of industrial inspectors which were published by the Royal Ministry of Commerce (*A Magyar Királyi Iparfelügyelők tevékenysége az 19XX. évben*). Industrial inspectors were required to visit plants with more than twenty regular employees or using at least one engine to check if security requirements were met since 1893. Their collected data on total engine power capacity[20] and employment by sector of manufacturing were published at the industrial inspector district level starting from 1901. I discuss some issues (change in the number of districts, non-complete coverage of assigned factories, treatment of outliers, etc.) with this data set in Appendix A.1.2.

I also digitized the balance sheets of Hungarian public limited companies in 1899 and 1912 which were published in the 1900 and 1913 editions of the Great Hungarian Compass (*Nagy Magyar Compass*, formerly known as *Mihók-féle Magyar Compass*). This data set provides a panel of book values of capital stock, equity and total assets, which are then aggregated at the county level. These data allow me to analyze the relationship between book capital and engine power capacity because many firms reported their engine power capacity as well. Further description of this data set (what I include in the book capital and equity measures, how the balance sheet of multi-plant firms is split across counties, etc.) can be found in Appendix A.1.3.

[20]Engine power is measured in horsepower. The performance of electricity-, gas-, steam- and water-driven and internal combustion engines is summed in the analysis.

1.3.5 Other data sources

The main additional data sources are censuses conducted in 1881, 1891, 1900 and 1910.[21] They provide data on the share of different ethnic groups (defined by mother tongue), religion, literacy rates and sectoral (primary and other sectors) employment. Agricultural characteristics (share of forests, agricultural mechanization, etc.) come from the agricultural survey of Hungary in 1895 (*A Magyar Korona országainak mezőgazdasági statisztikája*). Data on the share of ploughing land which was not sellable (mainly properties held in fee tail[22]) are published in the Hungarian Statistical Yearbook for 1893. Information on the financial sector can be found in the 35th volume of the Hungarian Statistical Bulletin (*A Magyar Szent Korona országainak hitelintézetei az 1894–1909. évben*). Data on retail prices, public goods provision (telegraph offices, length of roads and railroads, number of doctors and hospital beds) and students in compulsory primary education come from the Hungarian Statistical Yearbooks. Data on births and deaths are from the 46th volume of the Hungarian Statistical Bulletin (*A Magyar Szent Korona országainak 1901-1910. évi népmozgalma*). Trade data are from the 63rd volume of the Hungarian Statistical Bulletin (*A Magyar Szent Korona országainak 1882-1913. évi külkereskedelmi forgalma*).

1.3.6 Sample construction

County-level sample The highest level of aggregation used in the main analysis is county.[23] Above the county level, there were no additional layers of regional government - only the national level. The median county had a population of 220,000 in 1900. I exclude from the estimation sample the capital, Budapest, which would be an influential outlier (around one-quarter of all factories were found in the capital in 1900), so that I can focus on the industrialization of the countryside. Moreover, I omit Fiume/Rijeka (owing to its special political status as the sole sea port belonging to the Hungarian Crown, even though it was geographically located in the autonomous Kingdom of Croatia-Slavonia - see Figure A2) and Sáros county, the origin of the first documented emigrants. After these restrictions, I am left with 62 counties containing approximately ninety-five percent of the total population. County-level summary statistics are presented in Table A10.

County-level statistics suggest rapid development in Hungary at the turn of the

[21] The 1891 census has 01/01/1891 as reference date, while the 1900 and 1910 ones were supposed to record data pertaining to December 31st.

[22] Fee tail restricted the sale of real estate property, prevented it from being sold and caused it to pass automatically to an heir determined by a settlement deed. Many church properties could not be sold either because they were held in mortmain.

[23] I treat cities with royal free city rights as part of the county which surrounded them.

century. Between 1891 and 1900, county population grew by 9.3% on average. The rate of literacy or the number of industrial workers was rising quickly as well, the latter reaching 15% of all workers by 1900. Almost one-quarter of these industrial workers earned their living in factories. The rest were craftsmen or employed in smaller workshops. In addition, half of the increase in industrial employment stemmed from increases in factory employment around 1900. Despite ongoing industrialization, the GDP per capita of Hungary fluctuated around one half of that of Germany and was similar to that of Finland, Italy or Spain in the decades preceding WWI (Klein et al., 2017; Bolt and van Zanden, 2020), and the share of industrial employment was almost three (two) times larger in Germany (Austria) even in 1910 (Schulze, 2007a).

Municipality-industry subgroup level sample This sample consists of every municipality-industry subgroup pair which had at least one factory both in 1900 and 1910. I am left with 920 observations in 102 different industry subgroups. To classify the diverse subsectors, I use the industry classification of the Hungarian Royal Central Statistical Office throughout the entire analysis which can be found in Appendix A.1.1.[24] I do not exclusively consider manufacturing but also include mines, construction, touristic and utility firms in the sample. Industry subgroups employing the most workers and being present in the most municipalities in this panel can be found in Table A11. Besides sectors producing for export (mills, sawmilling, sugar production), coal mining, iron and steel production, spinning and weaving played a large role. The five largest industry subgroup-level employment expansions and contractions - in the full sample as well as in the panel - are presented in Table A12. The characteristic sectors of the Second Industrial Revolution (coal mining, machinery, steel-making) experienced rapid expansion and some construction-related sectors (brick production or sawmilling) also signal the growth in Hungary. The single largest drop in sectoral employment is related to the decline of medieval gold and silver mines.

Factories in the panel employed 89% of all factory workers in 1900 (approx. 195,000 workers). This share diminished to 73% by 1910, even though employment in the panel itself increased by more than 60,000 workers. Additionally, Table 1.2 shows that the median-sized municipality-industry subgroup cell employed 74 workers in 1900. This number increased to 115 over a decade. Furthermore, more than three-quarters of municipality-industry subgroup cells contained only a single factory. This fact ensures that the analysis of this data set is a close approximation to an establishment-level analysis.

[24]Published in the second volume of the 1900 census [p. 1-8]. Some industry subgroups were further disaggregated in the 1910 census, but I group those together to be consistent with the original classification in 1900.

Table 1.2: Summary statistics - municipality-industry subgroup sample (1900-1910)

	Mean	25th percentile	Median	75th percentile	90th percentile	S.D.	Total
Employment (1900)	189	37	74	182	422	358	173,920
Employment (1910)	262	53	115	290	605	467	240,610
Number of factories (1900)	1	1	1	1	2	1	1,189
Number of factories (1910)	1	1	1	1	2	1	1,266
N	920						

Note: own calculations.

1.4 Empirical strategy

In the first part of this section, I discuss the appropriate regression specification and the potential biases of an OLS estimation. Then, I propose an instrumental variable and present numerous empirical tests to support the validity of the exclusion restriction.

1.4.1 Regression specification

I assume the following functional form for the county-level regression:

$$\frac{\Delta\ y_{c,t}}{Population_{c,t}} = \beta \cdot \frac{\Delta\ Emigrants_{c,t}}{Population_{c,t}} + \gamma \cdot x'_c + \Delta\ \epsilon_{c,t}, \tag{1.1}$$

where $\Delta\ y_{c,t}$ represents the difference in an outcome variable (e.g., number of factory employees) in county *c* between year *t* and *t+10*. The change in the number of emigrants is defined as the difference between the number of locals who left for the US between *t* and *t+10* and the number of those who returned in the same period. Both the outcome variable and the independent variable of interest are divided by initial county population. In this way, $\frac{\Delta\ Emigrants_{c,t}}{Population_{c,t}}$ can be interpreted as net population loss as a share of initial population and β as the effect of one emigrated local to the US. X_c is a vector of controls, which is included in order to account for the potential growth effect of some local fundamentals (e.g., proximity to the capital). $\Delta\ \epsilon_{c,t}$ is the error term potentially correlated with the variable of interest. Finally, I include statistical region-fixed effects, $f_{r(c)}$, in a few, more restrictive specifications which control for region-time varying effects in this differenced specification.[25] They guarantee that region-specific simultaneous shocks cannot drive the results.

The municipality-industry subgroup panel allows me to control for additional potentially confounding effects:

$$\Delta\ y_{i,j,c(j),t} = \beta \cdot \frac{\Delta\ Emigrants_{c(j),t}}{Population_{c(j),t}} + f_i + \gamma \cdot x'_{c(j)} + \Delta\ \epsilon_{i,j,c(j),t} \tag{1.2}$$

[25] I use the seven statistical regions defined by the HRCSO.

where $\Delta y_{i,j,c(j),t}$ represents a change in a measure of industrial employment (e.g., log of factory employees) in sector *i*, municipality *j* and county *c* between year *t* and *t+10*. Sector-fixed effects (industry group or subgroup), f_i, are included to control for sector-specific, country-level growth shocks. This could be important if heavily emigration-exposed areas had different sectoral composition compared to less exposed regions and sectors experienced diverse growth shocks.

Nevertheless, running regression 1.1 or 1.2 may not identify the true β due to a number of issues. Reverse causality might be a problem as negative shocks to industrial growth can stimulate people to leave. This would lead to a downward bias in the estimated β. Omitted variable bias (OVB) can also plague the analysis. I interpret OVB in its broadest sense, testing for the potential time-varying effect of many fundamentals. For instance, the level of local human capital might not exclusively have a *level* but also a *growth* effect or formerly unexploited endowments could dynamically gain importance. If the latter two effects are correlated with emigration exposure, the estimate for β will be inconsistent. While time-varying fundamentals could be important in theory, this analysis exploits the decades of the Second Industrial Revolution. This lessens concerns over this type of confounders since local endowments which spurred growth in the 1890s (e.g., availability of coal) were most likely very similar to those in the 1900s. Last, HRCSO (1918) highlights that the administration of emigrants, especially of return migrants, might have contained some measurement error.[26] To the extent that it creates purely random noise, it can attenuate the estimated coefficient.

To mitigate the aforementioned concerns, I use an instrumental variable (IV). I demonstrate that the proposed IV does not predict industrial growth prior to mass migration. Consequently, pioneering emigrants did not leave steadily declining regions. To confirm that OVB does not contaminate the estimates, I present regressions showing that the IV is uncorrelated with many predetermined potential drivers of growth. An instrumental variable can also help in reducing the attenuation bias caused by measurement error in the number of migrants.

1.4.2 Instrumental variable

I use distance to the county from where the first pioneering emigrants left for the US as the baseline IV for the emigration exposure of a given county. The idea behind this instrumental variable can be summarized as follows. Based on HRCSO (1918, p. 8* and 25*), Nagy (1983, p. 177) or Puskás (1983, p. 268), I argue that the very first emigrants left from Sáros county because of its geographic (and linguistic) proximity to earlier emigrants in the neighboring western Galicia

[26] I discuss that this measurement error was minor and mainly manifested itself for out-migrants to Austria and return migrants from Romania (Appendix A.1.4).

(currently in southern Poland but part of Austria-Hungary during the relevant time period).[27] Thus, where people started to leave the country was most likely quasi-exogenous to local economic conditions.[28] A person in a location closer to the first pioneers was more likely to have pioneers in her own social proximity. The role of these pioneers was especially large for the barely literate, poor, agricultural population. They could reduce information frictions about the outstanding earnings opportunity in the US and financially help the emigration of their friends or relatives. Consequently, areas with more pioneers in the US responded more intensively to the pull shock of the booming US economy in the period of interest. In short, the IV aims to capture the higher probability of chain migration, a widely noted phenomenon from historical emigration periods (e.g., Hatton and Williamson, 1994; Spitzer and Zimran, 2019) to contemporaneous ones (e.g., Munshi, 2020), closer to the first emigrants. I collect supporting pieces of evidence specifically on Austro-Hungarian chain migration in Appendix A.2.1.

Sáros, the county in which authorities first registered emigration to the US was located in northern Hungary. Its surrounding region was not an economically left-behind area as it had a close to median level of development as measured by GDP per capita in 1870 (Schulze, 2007b). The lack of systematically different fundamentals between regions more and less exposed to emigration and the role of pioneers was also observed by contemporary[29] and current[30] authors.

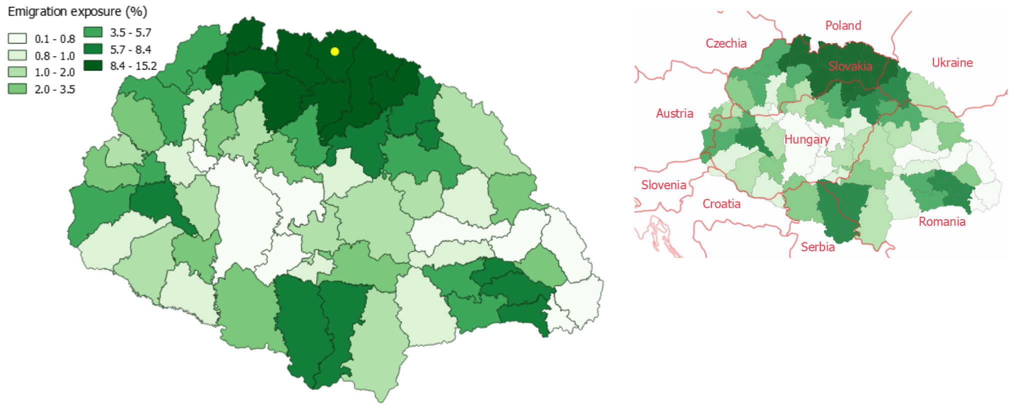
Figure 1.3a shows the spatial distribution of emigration exposure. There is a clear concentration of high-exposure counties around Sáros. While 83% of all emigrants left from the statistical region around Sáros in 1899, this ratio gradually dropped to 34% by 1905 and hit 21% by 1907. Afterwards, it stabilized around this level. HRCSO (1918) notes that subcounty units that were closer to the core emigration area near Sáros started to experience more intensive emigration to the US earlier. Nevertheless, some outlier counties in the south are worth highlighting. German minorities living in southern counties (Temes, Torontál), Western Transdanubia or Southern Transylvania exhibited higher emigration intensity (see Figure

[27] No railway lines crossed the northern Carpathian mountains towards Galicia until 1874 (Figure A7). There is evidence that even migrants from Galicia often chose the route via Vienna to reach northern German ports (Pálvölgyi, 2010).

[28] The lack of significant pre-trends on main outcomes in the decades of pioneering supports the assumption that permanent shocks to industrialization did not drive emigration (Table 1.3).

[29] HRCSO (1918, p. 47): «*comparing districts with similar intensity of emigration but rather different economic characteristics supports our claim that one of the main drivers of emigration are examples set by local pioneers.*» Éber (1902) emphasized that laborers faced essentially the same economic problems all over Hungary.

[30] Erdélyi Riport (2014): «*however, this simple description [more emigration-exposed areas lying far from urbanized centers and a population boom] does not explain why emigration happened in some regions, leaving other regions with similar structures untouched...if some families settled in the US from a given village, their relatives, neighbors could quickly follow them.*»



(a) Emigration exposure (US) at the county level (1900-1910)

(b) Pre-WWI Hungary and current country borders

Figure 1.3: Emigration exposure (US)

Note: own calculations. The yellow dot indicates Sáros county. Color categories are set so that each category contains an equal number of counties. The population-weighted raw correlation between proximity to Sáros (negative log-distance) and emigration exposure is 0.59. Red lines indicate current country borders.

A8). The reason behind this phenomenon was two-fold (HRCSO, 1918). First, locals were documented to be specifically targeted by agents of German shipping companies early on. Second, they also had close relationships with Germans in the German Empire which enabled them to receive information on emigration opportunities relatively early. These pioneers could set an example for non-Germans in their local network, too.

To exploit the contagious spread of emigration, I use the negative log-distance to Sáros county, that is, the proximity to Sáros, as an instrumental variable. I define the distance between two counties as the straight-line distance between their respective county seats (*megyeszékhely*).[31]

To enhance the causal interpretation of the estimates, I control for some potentially important confounding variables as well. Considering the distance-based nature of the IV, I include two correlated distance measures: proximity to Budapest (the capital of Hungary), which attracted masses of internal migrants as the engine

[31]The only exception being county Pest-Pilis-Solt-Kiskun where the county seat was Budapest. In this case, Kecskemét is set as the «county seat» which had a more central position in the county and was the only city there with royal free city rights apart from Budapest. The dense railway network implies that straight-line and travel distances are strongly positively correlated (see Figure A7). In the municipality-industry subgroup level panel, I measure distance between Sáros's county seat (/Budapest/Fiume) and *sub*county (*járás*) seats.

of growth, and proximity to Fiume/Rijeka, which was the most important seaport of Hungary at the Adriatic coast. Including proximity to this port accounts for differences in market access for maritime trade, but also controls for the fact that emigrants could leave through this port directly for New York starting in 1904. Additionally, I control for the share of industrial workers in all workers in 1891, a predetermined proxy for economic development. I refer to these three control variables as baseline controls in what follows. The identification assumption is that, conditional on these baseline controls, distance to Sáros affects second stage outcomes in a 2SLS regression only through its effect on spurring chain migration. The subsequent parts of this section provide corroborative evidence on the validity of this exclusion restriction.

1.4.3 Correlation with observable determinants of growth

In Table A14, I show that the proposed IV is not correlated with many predetermined,[32] growth-related variables in the preferred specification. The columns of Panel A show that the IV is not associated with previous population growth and density, the share of local-born population, broad agricultural characteristics or human capital-related measures. Variables in Panel B aim to capture financial development, ethnic characteristics, level of public goods provision and religiosity as well as inequality-related features of counties. Conditional on the baseline controls, the IV is not correlated with any of them.

1.4.4 Pre-trend analysis

Next, I lend additional support for the validity of the exclusion restriction by showing the absence of pre-trends on three of the most important outcomes. I use the following reduced form specification to do so:

$$\frac{\Delta\, y_{c,t}}{Population_{c,t}} = \beta \cdot Proximity\ to\ Sáros_c + \gamma \cdot Baseline\ controls'_c + \Delta\, \epsilon_{c,t}. \quad (1.3)$$

Pre-trend regressions are presented only for the county-level analysis in Table 1.3 because the list of factories (municipality-industry subgroup pairs) was not published in the 1891 census. The outcome variable in Columns 1-3 is the change in the number of industrial workers as a share of initial population in each period. Individual craftsmen as well as (factory) employees are included. Columns 1 and 2

[32]Outcome variables are measured in 1900 if not noted otherwise in the footnote of the table.

Table 1.3: Pre-trends and the period of interest - main outcomes

	Total industrial employment			Factory employment		# of factories		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1881-1891	1891-1900	1900-1910	1891-1900	1900-1910	1891-1900	1900-1910	1900-1910 (no min. & smelting)
Proximity to Sáros	-0.116	-0.030	-0.863**	0.231	-0.520***	-0.001	-0.037***	-0.037***
	(0.235)	(0.262)	(0.373)	(0.261)	(0.190)	(0.006)	(0.011)	(0.011)
Mean of outcome	0.63	1.23	1.66	0.60	0.85	0.04	0.06	0.06
Standard deviation of outcome	0.66	0.89	1.22	0.77	0.82	0.03	0.05	0.04
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	No	No	No	No	No	No
Sample size	62	62	62	62	62	62	62	62

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The outcome variable is the change in the number of industrial workers (Columns 1-3), of factory employees (Columns 4-5) and of factories (Columns 6-9), divided by county population in the initial year of every time period and multiplied by 100 (1000 in Columns 6-8). Workers in industry classes II/A and II/B/a-c are included. A slightly different definition of industry must be used between 1881 and 1891 (see Appendix A.1.1 for more details). Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

show the lack of any significant pre-trend for this variable. The coefficient in the period of interest is several times larger and significant.

The reduced form effect, focusing only on the number of factory workers, is shown in Columns 4 and 5. Column 4 shows a highly insignificant, positive pre-trend. This is reassuring since the main concern about the identification strategy is the possibility that pioneering emigrants left steadily declining regions. As shown in the table, the coefficient in the period of interest is more than two times larger in magnitude and highly significant. The reduced form impact on the number of factories is presented in Columns 6-8. I use all sectors of industry other than 'mining and smelting' in Columns 6 and 8.[33] Column 7 reports the baseline estimate in the period of interest, which also includes 'mining and smelting'. Reassuringly, there is a level of magnitude difference between the coefficient in the pre-period and the period of interest.

The previous empirical exercises strongly support the validity of the exclusion restriction; even so, I discuss some additional threats to identification in Appendix A.2.2 at length. I argue that demand or labor market spillovers from Galicia are unlikely to drive the results. The possibility of a confounding trade shock or border effects is also ruled out.

[33] As there are no data on establishments employing more than twenty employees in 'mining and smelting' in 1891, the total number of employees is used. This is an innocuous assumption since the overwhelming majority of employees (97%) in mining and smelting worked in establishments with more than twenty employees. However, only 202 out of 653 establishments in mining and smelting employed more than twenty employees in 1900. Therefore, I present separate coefficients for the change in the number of factories. Source: 5th volume of the 1900 census, p. 98.

1.5 Main results

This section starts with the first stage estimation and, then, presents results on population and sectoral employment growth. Changes in employment in factories are investigated both at the county and the municipality-industry subgroup level.

1.5.1 First stage

The first stage regression of the county-level estimation is shown in Table 1.4. Column 1 shows the regression using only the instrumental variable as an explanatory variable. Then, the two distance-based controls and the proxy for industrial development are added. The coefficient of interest does not change sign and somewhat shifts away from zero. Combined with a considerable increase in the R^2, this minor change in the coefficient suggests that the first stage is not driven by unobserved confounders (Oster, 2019).[34] The coefficient of interest can be interpreted as a semi-elasticity: a 1% reduction in proximity to Sáros predicts a 0.053 percentage point higher emigration exposure in the preferred specification. Including controls for region-time fixed effects only moderately diminishes the point estimate. Additionally, all estimated coefficients on the baseline controls have the expected sign. Proximity to Budapest, the booming capital, is correlated with less emigration. On the other hand, being close to Fiume/Rijeka, from where a direct shipping line was established to the US in 1904, predicts higher emigration rates. Also, industrial development does not seem to have a clear relationship with emigration, in line with the discussion in Section 1.4.2. Figure A9 additionally shows that the first stage relationship is indeed approximately linear.[35]

1.5.2 County-level estimation

1.5.2.1 Total population and broad sectoral employment

Before delving into the analysis of structural change, I explore the effect of emigration on total population at the county level in Table 1.5. I control for the lag of population growth since the strong autocorrelation of the process allows for a more precise estimation with a negligible change in the point estimate. The first thing to notice is that the 2SLS coefficient is slightly more negative than the OLS one. Controlling for region-time fixed effects makes the coefficient even more negative but more imprecisely estimated. The interpretation of Column 3

[34] I formally assess its sensitivity in Appendix A.2.3.

[35] The first stage regression of the municipality-industry subgroup sample is shown in Table A15. The coefficients are remarkably similar and the inclusion of industry group-fixed effects barely changes them.

Table 1.4: First stage - county-level analysis

	(1)	(2)	(3)	(4)
Proximity to Sáros	3.87***	5.43***	5.34***	4.51***
	(0.82)	(0.72)	(0.75)	(1.29)
Proximity to Budapest		-3.57***	-3.64***	-3.26***
		(0.72)	(0.70)	(0.91)
Proximity to Fiume/Rijeka		4.37***	4.06**	2.51
		(1.42)	(1.54)	(3.40)
% of industrial emp. (1891)			0.07	-0.08
			(0.10)	(0.11)
Mean of outcome	3.50	3.50	3.50	3.50
Standard deviation of outcome	3.44	3.44	3.44	3.44
Region FE	No	No	No	Yes
Kleibergen-Paap F-statistic	22.52	56.98	50.40	12.17
Sample size	62	62	62	62
R^2	0.34	0.49	0.49	0.62

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The outcome variable is emigration exposure (US) which is defined in the main text. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

is that the local population dropped by 41 people following the departure of one hundred emigrants to the United States. Alternatively, a one percentage point increase in emigration exposure led to a 0.41 percentage point lower population growth. To understand the causes behind this significantly smaller than one-to-one population decline, I decompose population growth into births, deaths, internal and international migration, and analyze these components in depth in Appendix A.2.4. I demonstrate that emigration to the US was substituted for other international directions (e.g., Austria, Germany, Romania) and that it also relatively slowed down internal out-migration principally to Budapest. In other words, some people would have migrated even in the absence of the US option but to other countries or to Budapest. I also find an imprecisely estimated positive effect on births, which is consistent, for instance, with the positive effect of stunted structural change and initially increasing wages (both effects documented later) on fertility or a Malthusian response to a negative population shock.

The census divides population into two categories: workers (*kereső*) and their dependents (*eltartott*).[36] Out of the 41 people lost for every one hundred emigrants, 28 were workers and the rest were dependents. Note that the dependent losses might not simply be the consequence of the emigration of housewives or children.

[36] I make one change with respect to the original classification. I reclassify supporting family members (*segítő családtag*; mostly children and wives) in agriculture as dependents because their labor supply was most likely quite different compared to full-time agricultural day laborers or servants.

Table 1.5: Population and worker losses (1900-1910)

	Population (workers + dependents) growth				Workers		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	2SLS	2SLS	2SLS	OLS	2SLS	2SLS
Emigration exposure (US)	-0.368***	-0.372	-0.408***	-0.680***	-0.270***	-0.281**	-0.413**
	(0.083)	(0.266)	(0.135)	(0.240)	(0.071)	(0.112)	(0.206)
Population growth (1891-1900; %)	0.903***		0.893***	0.794***			
	(0.159)		(0.135)	(0.139)			
Mean of outcome	7.96	7.96	7.96	7.96	2.43	2.43	2.43
Standard deviation of outcome	5.27	5.27	5.27	5.27	2.28	2.28	2.28
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	No	Yes	No	No	Yes
Kleibergen-Paap F-statistic		50.40	50.50	9.31		50.40	12.17
Sample size	62	62	62	62	62	62	62

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The outcome variable is population growth defined in percentages in Columns 1-3, and the change in the number of workers as % of county population in 1900 (Columns 4-6). Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

There is anecdotal evidence - see the following paragraphs - which suggests that the extensive margin of the labor market played a role, too. In particular, it is likely that wives and children would have had to work in the absence of adult (male) breadwinners in their family, or simply that they worked outside their home more often because emigration opened up employment opportunities as substitutes for emigrated agricultural workers. This implies a dampening effect on relative worker losses but a larger loss of dependents in response to emigration.

Next, worker losses are decomposed into three broad sectors: primary production (mainly farming but including fishery and forestry), industry and other sectors (day laborers not associated with primary production or industry, merchants, public servants, servants, etc.). This decomposition in Table 1.6 reveals some intriguing findings.

First, the insignificant coefficient on agricultural worker losses is striking if we consider the agricultural background of most emigrants. To gain a deeper insight, primary production worker losses are split by gender in Columns 5-6. The coefficient for males is negative and significant, though small in magnitude. It is insignificant and even positive for women instead of the expected negative effect caused by the emigration of women. My interpretation is that women and children started to perform more agricultural tasks and, consequently, they were classified as workers rather than dependents in the census. This interpretation is supported by anecdotal evidence on women and children starting to participate in heavier agricultural tasks in areas experiencing emigration (Neményi, 1911). It is also consistent with Dinopoulos and Zhao (2007) or Antman (2011) who show an increased incidence of child labor following low-skilled emigration, and with a growing literature demonstrating increased female labor in unpaid family work or subsistence work in families with emigrated members (Binzel and Assaad,

Table 1.6: Decomposition of worker losses (2SLS; 1900-1910)

	(1) = (2)+(3)+(4)				(2) = (5)+(6)	
	(1)	(2)	(3)	(4)	(5)	(6)
	Workers	Primary prod.	Industry	Other	Primary prod. male	Primary prod. female
Emigration exposure (US)	-0.293***	-0.065	-0.167***	-0.061**	-0.087***	0.023
	(0.067)	(0.045)	(0.054)	(0.025)	(0.033)	(0.043)
Population growth (1891-1900; %)	0.307***	0.056*	0.142***	0.109***	0.072***	-0.015
	(0.058)	(0.033)	(0.053)	(0.019)	(0.020)	(0.027)
Mean of outcome	2.43	-0.15	1.66	0.92	0.53	-0.68
Standard deviation of outcome	2.28	1.15	1.22	0.74	0.92	1.05
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	No	No	No	No
Kleibergen-Paap F-statistic	50.50	50.50	50.50	50.50	50.50	50.50
Sample size	62	62	62	62	62	62

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The outcome variable is the change in the number of workers (Column 1), of primary production workers (Column 2), of industrial workers (Column 3), of workers in other sectors (Column 4), of male primary production workers (Column 5) and of female primary production workers (Column 6) as % of county population in 1900. Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2011; Mendola and Carletto, 2012; Lenoël and David, 2019). Nonetheless, these coefficients must be interpreted cautiously because even contemporary statisticians, who composed the attached analyses for census volumes, stressed the difficulty of classifying women and children as dependent or non-dependent, particularly in farming.

The second question which Column 3 of Table 1.6 poses is whether the losses in industry were larger or smaller than the direct effect resulting from the emigration of *industrial* workers themselves. To answer this question, the first stage coefficient is split into industrial and non-industrial emigration exposure in Table 1.7. The instrument predicts that around 7.5% of emigrants had been employed in industry. In the second stage, the change in the number of industrial employees is the outcome variable, while the explanatory variable of interest is industrial emigration exposure. The coefficient in Column 4 of Table 1.7 combines two previous estimates: a county lost 17 industrial workers following the departure of one hundred emigrants and approximately 7-8 people out of one hundred emigrants were employed in industry before their emigration. Thus, a county lost on average 2.3 industrial workers for every single emigrated industrial worker. To estimate this coefficient more precisely, population growth in the pre-period is included as a control in Column 5. In this specification, the null hypothesis that the coefficient of interest equals -1 is rejected in favor of the alternative hypothesis ($\beta < -1$) at the 10% significance level. The conclusion is that the identified negative effect on industrial employment was significantly larger than the direct effect implied by the emigration of industrial workers, suggesting that emigration stunted local structural change.

There are two potential ways in which the disproportionately large industrial

Table 1.7: Disproportionately large industrial worker losses (1900-1910)

	Emigration exposure [US; (1) = (2)+(3)]			Industrial workers	
	(1) Total	(2) Industrial	(3) Non-industrial	(4)	(5)
First stage:					
Proximity to Sáros	5.3386***	0.3993***	4.9393***		
	(0.7520)	(0.1004)	(0.7358)		
Second stage:					
Emigration exposure				-2.3157*	-2.4135**
(US; only industrial workers)				(1.2487)	(1.0206)
Population growth (1891-1900; %)					0.1803***
					(0.0688)
Mean of outcome	3.50	0.41	3.09	2.31	2.31
Standard deviation of outcome	3.44	0.41	3.14	1.48	1.48
Baseline controls	Yes	Yes	Yes	Yes	Yes
Kleibergen-Paap F-statistic				15.82	15.35
$H_0 : \beta = -1$; $H_1 : \beta < -1$ (p-value)				> 0.1	0.09
Sample size	62	62	62	62	62
Share of predicted emigrants (%)	100.0	7.5	92.5		

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The outcome variable in Column 1 is emigration exposure, which is decomposed into emigrants who had been employed in the industry (Column 2) and all other emigrants (Column 3). The dependent variable in Columns 4 and 5 is the change in the number of industrial workers (1900-1910), as % of county population in 1900. However, in emigration statistics, industrial workers do not exclusively include industry classes II/A and II/B/a-c but also II/C (commerce and finance) and II/D (employees of railway companies, post offices, shipping companies, etc.). Therefore, for consistency, the change in the number of workers in classes II/C-D is also incorporated in the outcome variable in Columns 4 and 5. The affected two classes constituted a marginally small share of emigrants, so their inclusion does not matter for the explanatory variable of interest (HRCSO, 1918, p. 34*). If exclusively industry classes II/A and II/B/a-c were included in the outcome variable in Column 5, the coefficient would barely change from -2.41 (1.02) to -2.23 (0.89). In addition, Table A16 presents results under the conservative assumption that none of the return migrants took up an industrial job (see the details of the imputation process of the number of industrial emigrants in Section 1.3). The dependent variable of Column 2 is referred to as 'Emigration exposure (US; only industrial workers)' in the lower part of this table. Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

decline could happen. Industrial workers may have been displaced and reallocated to primary production following plant or workshop closures. Alternatively, some locals might have been stuck in primary production as no entering or expanding incumbent firms hired them. I show below that the latter channel played the dominant role and contributed to diminished relative agricultural worker losses, together with the extensive margin response of wives and children in farming.

1.5.2.2 Factory employment - county level

In this work, I focus mostly on factories as their emergence and expansion signalled economic growth and structural change in the heyday of the Second Industrial Revolution. As Chandler (1994, p. 3) put it: «*the modern industrial enterprise played the most fundamental role in the transformation of the West-*

Table 1.8: Factory employment losses (1900-1910)

	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS	(5) 2SLS	(6) 2SLS	(7) 2SLS-pop. share
Emigration exposure (US)	-0.0662***	-0.0975***	-0.0982***	-0.1237***	-0.1408*	-0.1527**	-0.0816***
	(0.0242)	(0.0357)	(0.0307)	(0.0341)	(0.0799)	(0.0680)	(0.0303)
$\frac{\Delta_{1900,1891}\ Factory\ employment}{Population\ (1891)}\%$			0.6774***	0.6063***		0.5539***	
			(0.1139)	(0.1384)		(0.1586)	
Mean of outcome	0.85	0.85	0.85	0.85	0.85	0.85	0.69
Standard deviation of outcome	0.82	0.82	0.82	0.82	0.82	0.82	0.67
Baseline controls	Yes	Yes	No	Yes	Yes	Yes	Yes
Region FE	No	No	No	No	Yes	Yes	No
Kleibergen-Paap F-statistic		50.40	25.10	49.73	12.17	10.98	50.40
Sample size	62	62	62	62	62	62	62

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The outcome variable in Columns 1-6 is the change in the number of factory workers (as % of county population in 1900). The dependent variable in Column 7 is the change in the share of factory employees between 1900 and 1910 (as % of county population in the respective years). Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

ern economies...[and] brought the most rapid economic growth in the history of mankind». Recent research also confirms the enormous impact of mechanization and production methods in factories on labor productivity (Atack et al., 2020).

The first column of Table 1.8 shows the OLS regression for the change in the number of factory workers on emigration exposure. The 2SLS estimation strategy is applied in the next column which moderately increases the magnitude of the coefficient of interest. The point estimate implies that ten factory workers were lost for every one hundred emigrants to the US. Note that, at face value, this number is in itself larger than the average predicted number of *all* industrial workers in one hundred emigrants (7-8). It also means that almost two-thirds (=10/17; see Column 3 of Table 1.6) of the entire decline in industrial employment may be attributable to factory employment losses. Another interpretation of the coefficient is that a one standard deviation increase in emigration exposure led to a 0.4 (=0.098*3.44/0.82) s.d. decrease in the expansion of factory employment. Column 3 shows that exclusively controlling for the lag of the outcome variable delivers a very similar coefficient compared to using the baseline controls. The identified effect becomes more negative when controls for region-time fixed effects or the lagged outcome are included together with the baseline controls. This is reassuring as it suggests that the identified effect is not driven by regional growth shocks. Finally, I define the outcome variable as $\frac{Factory\ employment_{c,1910}}{Population_{c,1910}}\% - \frac{Factory\ employment_{c,1900}}{Population_{c,1900}}\%$ in Column 7. The interpretation is that a one percentage point increase in emigration exposure slowed down the rise in the share of factory workers by 0.08 percentage point. These severe factory employment losses are consistent with rapid industrial employment expansion in the US as a result of mass immigration in the same time period (Sequeira et al., 2020; Tabellini, 2020).

Following Dustmann and Glitz (2015), who stress the importance of firm creation and destruction in response to a labor supply shock, I decompose the identified negative effect into the entry of initially non-existent and exit of existent municipality-industry subgroup cells between 1900 and 1910, and employment changes in cells existing in both years. This analysis is a close approximation to an establishment-level analysis since the vast majority of cells contain a single factory (see Table 1.2). The first column of Table A17 suggests that roughly half of the factory employment losses stemmed from a slower expansion of municipality-industry subgroup pairs which existed both in 1900 and 1910. The remaining other half is exclusively attributable to a lower entrance of factories in formerly non-existent sectors. The coefficient of the exit channel is a tightly estimated zero. Thus, the decomposition suggests that factory owners sought to utilize their already installed capital instead of shutting down factories and laying off workers.

I present additional robustness checks in Appendix A.2.5 which include showing that the results are not driven by outliers in terms of industrial growth; that I do not capture a spurious North-South growth shock in the 1900s; and that the point estimate is similar when using a second instrumental variable (share of Germans) which also allows me to carry out overidentification tests. The high p-value of the latter bears out the validity of the overidentifying restrictions. Addressing additional concerns about statistical inference, I demonstrate that accounting for the potential spatial correlation of standard errors following Conley (1999) or bootstrapping them has a marginal effect on the p-value.

1.5.3 Factory employment - municipality-industry subgroups

Analyzing the effect on factory employment at the county level does not exploit the fact that the level of observation in the factory census is a municipality and narrowly-defined industrial sector pair. This disaggregated panel data set allows me to additionally control for sector-specific, time-varying effects, lending further support to the causal interpretation.

The results using the municipality-industry subgroup panel are presented in Table 1.9. The outcome variable is the difference in log-employment between 1900 and 1910, approximating employment growth. Each observation is a municipality-industry subgroup cell which had at least one factory in both census years. The OLS point estimate has the same sign as and is similar in size to the 2SLS one. More importantly, the 2SLS coefficient remains practically unchanged after the inclusion of industry group-fixed effects in Column 3, suggesting that the previously identified negative effect is not driven by a particular sectoral composition correlated with distance to Sáros and sector-specific growth shocks. The coefficient implies that a one percentage point increase in county-level emigration exposure led to a 2.2 percentage points lower sectoral employment growth locally.[37]

[37] I also implement nearest neighbor matching in Appendix A.2.5 which delivers similar findings

Table 1.9: Municipality-industry subgroup level employment growth (1900-1910)

	Baseline results				Robustness checks		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	2SLS	2SLS	2SLS	Fewer than 500 employees (1900)	No weighting	5+ locations
Emigration exposure (US)	-0.0044	-0.0207**	-0.0224***	-0.0602***	-0.0300***	-0.0246***	-0.0239**
	(0.0074)	(0.0096)	(0.0084)	(0.0230)	(0.0082)	(0.0060)	(0.0093)
Mean of outcome	0.15	0.15	0.15	0.15	0.26	0.35	0.15
Standard deviation of outcome	0.54	0.54	0.54	0.54	0.62	0.65	0.53
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry group FE	No	No	Yes	Yes	Yes	Yes	No
Industry subgroup FE	No	No	No	No	No	No	Yes
Region FE	No	No	No	Yes	No	No	No
Kleibergen-Paap F-statistic		116.32	119.14	22.51	134.94	125.53	133.38
Sample size	920	920	920	920	844	920	809
Number of clusters	60	60	60	60	60	60	60

Note: Unit of observation is municipality-industry subgroup. Robust standard errors, clustered at the county level, are in parentheses. All specifications (with the exception of Column 6) are weighted by the number of employees in 1900. The dependent variable is the growth of factory employment, defined as the difference in log-employment between 1900 and 1910. The dependent variable is winsorized at the 1st and 99th percentile. *Robustness checks:* the estimation includes observations with fewer than 500 workers in 1900 in Column 5. All observations are equally weighted in Column 6. The sample is restricted to industry subgroups with at least 5 observations in Column 7. Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Significance levels: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.6 Additional empirical results

So far, I have documented local deindustrialization in the wake of emigration. In this section, I present additional empirical findings to gain a deeper understanding of the underlying mechanism. Following Dustmann et al. (2008), I analyze the effect of emigration on the output mix, factor prices and production technology. More specifically, I discuss the effect on capital-/labor-intensive and more/less tradable sectors, low-skilled wages, mechanization and capital stock. I also address the evolution of basic goods' prices and the effect on municipalities at different stages of industrialization. The second stage coefficient is reported for specifications using the original 1900-1910 period, while the reduced form is shown for other time spans.

1.6.1 Output mix

Besides certain theoretical models (e.g., Heckscher-Ohlin), empirical research also motivates the analysis of local sectoral composition changes. Namely, it has been established that relative endowments played a crucial role during the process of industrialization within-country (e.g., Kim, 1999; Crafts and Mulatu, 2006; Martínez-Galarraga, 2012). Since most early factory laborers were initially farm workers (de Pleijt et al., 2020), a reduction in the local labor endowment owing to the emigration of agricultural laborers could have a particularly negative effect on labor-intensive industrial sectors. A measure of sectoral labor intensity is needed to test this hypothesis. I use the industrial inspector records (1901) to

to regressions.

calculate the engine power capacity-to-worker ratio for each industry group in the estimation sample. Sectors for which this ratio is above the median are classified as capital-intensive: iron and metal, chemical, food and paper industries, and machine manufacturing. All other branches of manufacturing (building materials, leather, lumber, textile sectors and printing) are labelled as labor-intensive.[38] Reassuringly, the balance sheets of public limited companies suggest that all industry groups classified as labor-intensive had lower book capital-to-worker ratios than any of the capital-intensive branches, which lends further support to the classification (see Appendix A.1.3). Unlike industrial inspector reports, the book capital-to-worker ratio allows me to show that mining was labor-intensive. This is plausible since (Hungarian) mining was technologically backward and excessively manual in this time period (Dix, 1988; Kaposi, 2002, p. 244). Jerome (1934) and Nikolić (2018) find a broadly similar labor intensity ranking for industrial sectors in the early-twentieth-century United States and in interwar Yugoslavia, respectively.[39,40]

The factory employment loss decomposition by labor intensity is reported in Table 1.10. The results are clear: three-quarters of the total effect stemmed from labor-intensive sectors. The coefficient on the change in capital-intensive employment is, if anything, negative as well, suggesting the absence of a rise of capital-intensive sectors that would be implied by the textbook Heckscher-Ohlin model. However, the small effect on capital-intensive sectors is not a result of relative sector size: approx. 115,000 and 80,000 factory workers were employed in labor- and capital-intensive sectors in 1900, respectively. To examine the robustness of these findings, I do a similar sample split in the municipality-industry subgroup level panel. The results in Table A18 are fully in line with the county-level analysis.

Yet analyzing the change in output mix by labor intensity is not the only relevant decomposition. The loss of local demand due to emigration could have serious repercussions on local industrial growth as well. Sectors relying more on local buyers might have experienced a more pronounced deceleration compared to sectors producing more traded goods. To test this competing hypothesis, I calculate the spatial concentration index of Ellison and Glaeser (1997) for every industry subgroup

[38] As an alternative measure, I calculate the same ratio for public limited companies in 1913, which corroborates the baseline classification (see Appendix A.1.3). In line with industrial inspector reports, this also indicates that the relative labor intensity of sectors did not change much over time, meanwhile the absolute level of the engine power-to-worker ratio increased in all sectors, suggesting widespread capital deepening.

[39] Jerome (1934) reports the share of wage bill in value added in 1925 or the horsepower per wage earner ratio in 1899. Nikolić (2018) calculates the output share of unskilled labor.

[40] None of the two measures of labor intensity contain construction and hospitality. I classify construction as labor-intensive considering its rather manual nature historically. Hospitality is classified as labor-intensive as well. Dropping construction and hospitality in Columns 2 and 5 of Table 1.10 would leave the coefficients practically unchanged.

Table 1.10: Decomposition of factory employment losses by labor intensity (1900-1910)

	(1) = (2)+(3)			(4) = (5)+(6)		
	(1) Total	(2) Labor intensive	(3) Capital intensive	(4) Total	(5) Labor intensive	(6) Capital intensive
Emigration exposure (US)	-0.0975***	-0.0742***	-0.0233	-0.1237***	-0.0949***	-0.0288*
	(0.0357)	(0.0273)	(0.0177)	(0.0341)	(0.0281)	(0.0160)
$\frac{\Delta_{1900,1891}\ Factory\ employment}{Population\ (1891)}\%$				0.6063***	0.4786***	0.1277**
				(0.1384)	(0.1641)	(0.0640)
Mean of outcome	0.85	0.54	0.31	0.85	0.54	0.31
Standard deviation of outcome	0.82	0.62	0.37	0.82	0.62	0.37
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	No	No	No	No
Kleibergen-Paap F-statistic	50.40	50.40	50.40	49.73	49.73	49.73
Sample size	62	62	62	62	62	62

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The outcome variable is the change in the number of factory workers between 1900 and 1910 (as % of county population in 1900). Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Labor-intensive industry groups are mining and smelting, the building materials, lumber, leather and textile (spinning and weaving as well as clothing) industries, construction, printing and hospitality. Capital-intensive industry groups are the iron and metal, paper, food and chemical industries, and machinery. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

which was present in at least two municipalities in 1900.[41] Total employment in factories at the county level is chosen as a proxy for market size following Ellison and Glaeser (1997).[42] The higher the value of the index, the more spatially concentrated a given industry subgroup was as a consequence of spillovers or natural advantages. Low index values indicate, meanwhile, that the spatial distribution of a given sector and of the whole economy roughly coincided. Furthermore, proximity to consumers matters relatively more, and natural advantages and spillovers less for these "low-index" branches of industry. Using the calculated spatial concentration index, I split the sample into localized and dispersed industry subgroups. I call a sector localized if its index value is above the sample median (vice-versa for dispersed sectors). The results of this decomposition exercise are shown in Table A19 and suggest that there are no significant differences across sectors based on their spatial concentration. Instead, factory employment significantly declined both in localized and dispersed sectors, and if anything, the point estimate on more traded (localized) sectors is larger. Hence, missing local demand does not seem

[41] Mian and Sufi (2014) and Gervais and Jensen (2019) also exploit spatial concentration to proxy for tradability.

[42] Unlike the construction of the original index, my index is estimated at the municipality-industry subgroup rather than the firm level. This is not a large deviation if we consider that most municipalities had a single factory in an industry subgroup (see Table 1.2). The results are very similar when the proxy for market size is county population in 1900, or when exclusively industry subgroups that were present in at least five municipalities in 1900 are used. These results are available upon request.

to explain the results. This conclusion is somewhat expected if we consider that the consumption of low-skilled laborers mainly consisted of food and beverages (66% of consumption), housing (10%) and clothing (9%; Cvrcek, 2013). If the demand channel had played a large role, then emigration should have primarily led to the decline of the food and beverages industry, whereas this sector, which is capital-intensive, was only mildly affected.

1.6.2 Low-skilled wage growth

After establishing that sectors more exposed to labor cost changes shrank more severely, an analysis of low-skilled wage dynamics is essential. Estimates for the growth rate (defined as the difference in log-levels) of the nominal daily wage of adult male agricultural day laborers are presented for the high and low season (summer and winter) in Table 1.11. 1898 is set as the starting year when the United States left behind the Panic of 1896 and emigration started to increase rapidly. In addition to estimates for the whole time period until WWI, I also present results for 1900-1910 as a comparison with industrial average wage growth and earlier census-based findings.[43]

Estimates for the entire period suggest that the effect on the wage of male laborers was insignificantly negative. However, this insignificant negative coefficient hides highly significant but opposing wage effects in the short/medium and long run. Namely, agricultural wages increased significantly during the initial years of mass migration, but this rise was entirely reversed in the later period. A reasonable theoretical interpretation is that the decreasing marginal product of labor implied steadily increasing low-skilled wages in the early years, but changing production technologies, potentially biased towards a more intensive use of capital, or a weak growth in the local capital stock made this effect disappear over time. The finding on a short and medium-term wage increase aligns well with the majority of the literature which, mostly in modern contexts, finds overall positive wage effects in comparable time spans (Mishra, 2014). However, the reversal of relative wage gains is remarkable because it happened in spite of steady emigration, which tends to work in the opposite direction, raising wages. Therefore, there must have been a strong counteracting channel depressing wage growth in the longer run.

To conduct a subcounty level investigation, I digitized subcounty-level agricultural wages. This analysis is more informative since the subcounty level provides a better picture of local labor markets, particularly given the limited commuting opportunities in Austria-Hungary.[44] Reassuringly, the results in Table A20, using

[43] Additionally, I digitized the daily wages of laborers in the forty-three regional markets which were reported in statistical yearbooks in the 1880s. Leveraging this data set, I can show that proximity to Sáros does not predict wage growth between 1880 and 1890 (unreported).

[44] A subcounty was i) a town surrounded by villages, ii) a town with its own council, or iii) a city

Table 1.11: Agricultural and manufacturing wage growth (1898-1912)

	Agricultural (summer, male)					Agricultural (winter, male)	Manufacturing (male)
	(1) 1898-1912	(2) 1898-1912	(3) 1898-1903	(4) 1903-1912	(5) 1900-1910	(6) 1900-1910	(7) 1900-1910
Proximity to Sáros	-0.010	-0.079	0.080***	-0.082***	-0.037	-0.022	-0.020
	(0.030)	(0.056)	(0.015)	(0.026)	(0.030)	(0.028)	(0.033)
Mean of outcome	0.62	0.62	0.01	0.61	0.45	0.47	0.55
Standard deviation of outcome	0.10	0.10	0.08	0.10	0.09	0.10	0.14
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	Yes	No	No	No	No	No
Sample size	62	62	62	62	62	62	62

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The outcome variable is the difference in the log-wage of adult male agricultural day laborers (summer: Columns 1-5; winter: Column 6) and of weekly-paid male manufacturing laborers (Column 7) in the indicated time intervals. All dependent variables are winsorized at the 10th and 90th percentile. Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

more than 500 subcounty units, are both qualitatively and quantitatively similar to those in Table 1.11.

A related question is the skill composition of agricultural day laborers. Positive selection into emigration among low-skilled agricultural workers could lower the average quality in the pool of workers which, in turn, could explain the negative effect on wages. However, the instrumental variable does not predict differential changes in the adult literacy rate in my preferred specification. Digitizing Austro-Hungarian military records, I find that the share of missing draftees soared in areas exposed to emigration. Yet the correlation between the share of missing draftees and the share of draftees below minimum height (or rejected owing to physical weakness) is practically zero. The lack of positive selection into emigration is consistent with Abramitzky et al. (2012) who find no evidence of selection from rural areas but do find negative selection from urban areas in Norway in the Age of Mass Migration. It has also been established that networks of past emigrants can mitigate selection on wealth (David and Jarreau, 2017). As these networks grow, selection might change from being positive to negative (McKenzie and Rapoport, 2010; Spitzer and Zimran, 2018), potentially resulting in an overall neutral selection.[45]

with free city rights.

[45]Results on selection are available upon request. It might still be argued that selection into emigration was mainly determined by unobservables, such as individualistic personality traits (Knudsen, 2019), and only those locals stayed who were reluctant to work in industry. However, this argument is inconsistent with several empirical facts. First, this line of thought cannot explain why predominantly labor-intensive sectors declined, which likely did not require more skills relative to capital-intensive ones. Second, agricultural and industrial laborers earned similar compensations over time implying close substitutability, as I discuss in this section below. Finally, the loss of more «entrepreneurial» laborers may explain fewer craftsmen or small industrial workshops but not fewer factories. Had these talented individuals stayed, they would have hardly ever been able to secure

The reduced form effect on male average wage growth in manufacturing is presented in the last column. This is likely a lower bound on the negative effect since the wage survey in the 1900 census suggests that weekly wages in labor-intensive sectors were 10-20% lower.[46] Consequently, the effect on within-industry group wage growth might have been even more negative considering that emigration mostly stunted the expansion of labor-intensive sectors. That is, the changing industrial composition could dampen the negative effect on average wage through the inclusion of endogenously fewer laborers from labor-intensive sectors with lower wages in the average wage calculation.

The results also suggest that agricultural and industrial low-skilled workers were likely to be close substitutes given their comparable wage growth coefficients. This is not surprising since past research has shown that the Industrial Revolution caused «deskilling» or the hollowing out of the skill distribution, and large industrial employers required limited skills from laborers following workshop-to-factory shifts[47] (Katz and Margo, 2014a; de Pleijt and Weisdorf, 2017). Alston and Hatton (1991) observe that agricultural compensation was similar to that in manufacturing within geographic regions in the US prior to the Great Depression. Margo (2000) finds no difference in common laborers' wages between farm and nonfarm sectors on local labor markets in the Antebellum US either. Additionally, Hungarian emigrants found jobs mainly in mining or manufacturing in the US, even though they most often had an agricultural background. This observation supports the notion of close substitutability between agricultural and industrial laborers.

1.6.3 Employment and mechanization in manufacturing

The findings on industrial employment decline and the reversal of relative wage gains could be consistent with capital-intensive technology adoption substituting for laborers. Hence, examining the impact on mechanization is key to uncover the underlying economic mechanism.

To this end, I compiled a data set from reports of industrial inspectors, which contains data on engine power capacity and employment in factories for every industry group in manufacturing at the industrial inspector district level in 1901 and 1912. Thus, unlike the previous analyses, construction, hospitality and mining are not part of the sample. I use wild cluster bootstrapping with small sample correction to calculate p-values because there were only 15 industrial inspector

loans to establish a factory, which were normally founded by foreign or Budapest-based capital owners, and local aristocracy (see Section 1.6.4).

[46] This is consistent with Atack et al. (2004) who show that wages increased in capital intensity in the US manufacturing between 1850 an 1880.

[47] de Pleijt et al. (2020) advocate the use of farm-to-factory transition instead since the majority of factory workers were previously (unskilled) farm workers.

Table 1.12: Engine power capacity and employment growth (1901-1912)

	Engine power capacity growth			Employment growth		
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity to Sáros	-0.319	-0.311	-0.306	-0.214	-0.198	-0.196
Ln(employment in 1901)		-0.077	0.229		-0.254	-0.264
Ln(engine power capacity in 1901)			-0.378			0.011
Bootstrapped p-value ($H_1 : \beta_{Prox.\ to\ Saros} \neq 0$)	0.022	0.048	0.088	0.059	0.059	0.053
Bootstrapped p-value ($H_1 : \beta_{e.p.} < \beta_{emp.}$)	0.075	0.113	0.140	0.252	0.097	0.105
Below/above median weight coeff.	-0.596/-0.351	-0.613/-0.343	-0.414/-0.309	-0.397/-0.201	-0.321/-0.184	-0.322/-0.18
Mean of outcome	0.91	0.91	0.91	0.61	0.61	0.61
Standard deviation of outcome	0.55	0.55	0.55	0.43	0.43	0.43
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Industry group FE	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	122	122	122	122	122	122
Number of clusters	15	15	15	15	15	15

Note: Unit of observation is industrial inspector district-industry group. Bootstrapped p-values are generated by Wild cluster bootstrap (with 999 replications, Rademacher weights, small sample correction and clustering at the industrial inspector district (1901) level). The dependent variables are winsorized at the 10th and 90th percentile. Columns 1-3 (4-6) are weighted by sectoral engine power capacity (employment) in 1901. The outcome variable is the difference in the log of engine power capacity (measured in horsepower; Columns 1-3) and of employment (Columns 4-6) between 1901 and 1912. Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891.

districts in 1901 (not counting Budapest and Fiume/Rijeka).

Estimates for engine power capacity and employment growth are presented in Table 1.12.[48] Columns 1-3 show a significant negative reduced form effect on engine power capacity growth. In line with earlier results, the effect on employment growth is also significantly negative. I re-ran the regressions for observations below and above the median weight and report their point estimates in the same table. They suggest that the results are not driven by a handful of influential observations. The crucial finding is that the point estimate on engine power capacity growth is (borderline) significantly larger in absolute value than the effect on employment growth, strongly suggesting the lack of induced mechanization and that firms, in the long run, did not counteract wage increases with capital-intensive technology adoption.

1.6.4 The capital stock of public limited companies

Another way to shed light on the effect of emigration on the capital stock is examining the balance sheets of public limited companies (PLCs). Since Hungary was a laggard in industrialization, more factories were established by PLCs than in Western European countries, where factories often gradually grew out of privately-held family businesses or workshops (Kozári, 2009). Hence, approximately half of non-governmental factory employment was connected to PLCs according to the 1900 census. Their economic significance must have grown rapidly as the number

[48] I show that the instrument does not predict the log-engine power capacity or log-number of workers in 1901, irrespective of the weighting used (Table A22). This further corroborates the claim that areas closer to the pioneers were not differentially developed.

of PLC-owned plants increased almost three-fold in my sample by 1913.

The availability of the balance sheets of PLCs allows me to analyze the impact on the capital stock from a different angle (book value instead of engine power capacity) and with a higher effective sample size. As explained in Appendix A.1.3, I construct a measure of book capital (mainly the value of equipment, machines and factory building) and study the effect on its change at the county level, normalized by initial county population in 1900. In Table A23, I use my preferred specification and demonstrate that the reduced form effect on changes in the book capital stock is insignificantly negative. However, I identify a significant negative effect on labor-intensive sectors.[49] The effect on capital-intensive sectors is insignificant and turns negative after the inclusion of region-time fixed effects (not reported). I also estimate the effect on equity and total assets in labor-intensive sectors and find a significant negative coefficient for both variables. Taken together, these results corroborate the evidence on mechanization and establish capital losses in labor-intensive sectors for the case of public limited companies.

1.6.5 The evolution of regional prices

I digitized data on the price of staples to investigate how nominal wage changes relate to real wage changes. Even though the government collected the prices of basic goods in hundreds of markets, only average prices at the main regional markets were published.[50] Furthermore, the only market reported in the core area of emigration near Sáros is Kassa (*Košice*).

Products whose prices are presented in Table A13 represented close to two-thirds of the consumption basket of laborers (Cvrcek, 2013). The analysis of these prices reveals that Hungary was a well-integrated economy by 1900. The coefficient of variation calculated for the prices of easily traded goods (e.g., grains, sugar or wine) across markets was generally below ten percent. Additionally, a comparison of inflation rates between Kassa and other markets shows that price changes around Kassa followed the country's average, which is consistent with the historical narrative that regional prices closely followed the Budapest commodity exchange after the expansion of the telegraph system (Kaposi, 2002, p. 250; Gao and Lei, 2021). The absence of differential inflation across regions suggests that changes in the real wage were mostly driven by changes in the nominal wage rate across labor markets. Moreover, this evidence of integrated markets and limited spatial price differences also supports the finding that local demand effects due to emigration were modest.

[49] To be consistent with the analysis relying on industrial inspector records, mining is not included. The reduced form coefficient in Column 4 increases from -0.013 (0.007) to -0.016 (0.007) after the inclusion of that sector.

[50] See the 36th volume of the Hungarian Statistical Bulletin [p. 106].

Table 1.13: Decomposition of factory employment losses
by the local level of industrialization (1900-1910)

	(1) = (2)+(3)		
	(1) Total	(2) Industrialized	(3) Non-industrialized
Emigration exposure (US)	-0.0972***	-0.0221	-0.0752***
	(0.0357)	(0.0253)	(0.0244)
Mean of outcome	0.85	0.29	0.56
Standard deviation of outcome	0.82	0.64	0.46
Baseline controls	Yes	Yes	Yes
Region FE	No	No	No
Kleibergen-Paap F-statistic	50.40	50.40	50.40
Sample size	62	62	62

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The dependent variable is the change in the number of factory employees in all municipalities (Column 1), in industrialized municipalities (Column 2) and in non-industrialized municipalities (Column 3) - aggregated at the county level. All outcome variables are expressed as % of county population in 1900. The definition of industrialized municipality can be found in the main text. Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Székelyvarság municipality did not exist in 1900, so it is omitted from this decomposition. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.6.6 Heterogeneous effect by the initial level of industrialization in municipalities

The municipalities of pre-WWI Hungary stood at various stages of industrialization. The booming industry of cities like Győr, Pozsony (*Bratislava*) and Temesvár (*Timişoara*), and the rapidly growing network of smaller municipalities in major industrial regions (current central Slovakia or Caraş-Severin in Romania) prospered, whereas some parts of the country were still largely stuck in agricultural production. An important question is to understand whether emigration slowed down the gradual spread of industrialization to farming-dominated, rural areas or, perhaps, acted as a drag on growth mainly in industrial centers lacking cheap labor for new factories.

To study this, I calculate the share of industrial workers in the local population in 1900. I define every municipality which had a factory in 1900 or 1910 as industrialized if the local share of industrial workers was above the 75$^{\text{th}}$ percentile (approx. 17.5%; percentage of total population). According to this definition, around 82,000 factory workers were employed in non-industrialized and 114,000 in industrialized municipalities in 1900. Factory worker losses in these two types of locations are decomposed in Table 1.13. The results are easy to interpret: the overwhelming majority of the negative effect was concentrated in not yet industrialized locations. Therefore, emigration slowed down the industrialization of locations which were initially more agricultural, thus widening the inequality of development between the leaders and laggards of industrialization within Hungary.

1.7 Conclusion

Economists traditionally emphasize the role of high-skilled, well-educated individuals when emigration from developing countries is discussed (e.g., Docquier and Rapoport, 2012). In this work, I show that the emigration of lower-skilled locals may have serious repercussions as well. Combining various archival data sources, I demonstrate that the emigration of mostly agricultural laborers led to the slowdown of local industrialization in Hungary prior to WWI, curbing factory employment growth in labor-intensive industrial sectors in particular. Additionally, emigration did not induce mechanization in manufacturing and low-skilled wages failed to increase in the long run as well. Unfortunately, it is impossible to investigate the persistence of the identified negative effects empirically because of the dissolution of Austria-Hungary at the end of WWI. Nonetheless, looking at a longer time horizon might offer a silver lining to emigration-exposed areas. Atkin (2016), Bustos et al. (2019) and Franck and Galor (2021) show that the expansion of low-skilled-intensive industrial sectors may inflict long-term negative effects on human capital accumulation, manufacturing productivity and skilled-intensive technology adoption.

A Appendix - Chapter 1

A.1 Data appendix

A.1.1 Industrial classification

In this work, workers are classified as industrial workers if they were employed in industry classes II/A or II/B/a-c, according to the classification below.

Industry group classification (1900; subgroup codes in parentheses):

o Class II/A

II/A Mining and smelting (11-19)

o Class II/B/a

II/B/a/I Iron and metal industry (20-41)

II/B/a/II Machine and transport equipment manufacturing, electric power generation, musical and scientific instrument production (42-55)

II/B/a/III Building material production (56-66)

II/B/a/IV Lumber and bone industry (67-85)

II/B/a/V Leather, fur, feather and rubber industry (86-92)

II/B/a/VI Spinning and weaving (93-101)

II/B/a/VII Clothing industry (102-120)

II/B/a/VIII Paper industry (121-123)

II/B/a/IX Food and beverages industry (124-153)

II/B/a/X Chemical industry (154-169)

II/B/a/XI Construction (170-187)

II/B/a/XII Printing, photographer, engraver and painter (188-192)

II/B/a/XIII Hospitality (coffee houses, hotels, pubs, restaurants, spas, etc.) (192-198)

II/B/a/XIV Other industrial occupations (199)

- Class II/B/b

II/B/b Cottage industry (200)

- Class II/B/c

II/B/c Itinerant craftsmen (201)

More than 1 million workers were employed in industry classes II/A and II/B/a in 1900. The number of cottage industry workers (fewer than 40,000) and itinerant craftsmen (fewer than 7,000) was negligible in comparison.

Albeit the definition of 'Industry' (II/A and II/B) did not change between 1900 and 1910, earlier changes are important for pre-trend estimation. Before 1900, some types of gardening and fishery belonged to the industry (later to primary production), and lime/magnesite burning, quarrying, clay and sand mining to mining (later to II/B/a/III). Some services - e.g., carriers, hucksters, porters, undertakers - belonged to the industry, and railroad construction and water regulation works were classified as part of Traffic (II/D) instead of Construction (II/B/a/XI) before 1900. Fortunately, none of these reclassifications matters for the number of *total industrial workers* in the 1891-1900 period because the 1900 census provides adjusted figures for 1891, allowing for an accurate comparison between 1891 and 1900. However, data in the 1881 census cannot be adjusted to reflect the new definition of industry. Therefore, the 1881-1891 pre-trend estimation for total industrial employment uses the *old* definition of industry in both years.

The pre-trend of *factories* (1891-1900) is affected by some minor measurement problems, too. On the one hand, firms in hospitality employing more than twenty workers were not included in the 1891 county level numbers (their employment was less than 2,500 workers in 1900). On the other hand, big employers in gardening (2 firms with 44 employees) and fishery (1 firm with 35 employees) were included in 1891. The aforementioned reclassification of some services (e.g., hucksters) does not matter since workers in those sectors did not work in large establishments. The reclassifications in construction (e.g., railway construction firms from Traffic to Construction) has a tiny effect as the affected sectors jointly employed fewer than 3,000 workers in 'factories' in 1900. In conclusion, owing to the rather limited employment involved (the affected sectors did not employ jointly more than 3% of all factory workers in 1900), these classification changes should not affect the pre-trend results in any meaningful way. Reassuringly, the coefficient in Column 4 of Table 1.3 changes from 0.23 (0.26) to 0.22 (0.26) after excluding gardening-fishery in 1891 and all new industry subgroups in 1900 (not reported).[51]

Last, some industry subgroups would not be treated as narrowly-defined industry today. For example, power plants (fewer than 1,000 employees in 1900),[52] sanitation firms (671 workers in 1900) and water works (548 workers in 1900) would be classified as utilities, while barbers (17,574 workers in 1900), laundries (556 workers in 1900), skinners (1,133 workers in 1900) or washing and ironing (26,412 workers in 1900) would be considered services. Moreover, cattle or pig fattening (joint employment less than 1,000 in 1900) would be treated as agricultural activities. Fortunately, the role of such sectors was limited in the economy as measured by employment (less than 5% of the industrial total), so the overwhelming majority of industrial workers were employed in actual mining and manufacturing and, to some extent, in construction and hospitality. More importantly, the presence of these sectors among factories was almost non-existent, except for power plants. For a comparison of classifications, Federico and Klein (2010) include in their definition of industry mineral extraction together with manufacturing, construction, and gas, electricity and water, following modern European convention.

[51]The only source of measurement difference left is the employment of some fewer-than-20-employee mines, quarrying sites and kilns included in 1891. The number of factories coefficient in Column 6 of Table 1.3 changes only mildly as well from -0.00142 (0.0062) to -0.00138 (0.0068). Here, industry subgroups 56, 57, 61 and 147 are subtracted from the number of factories in 1900 as they were classified as mining in 1891.

[52]Budapest is also included in this paragraph.

A.1.2 Industrial inspector reports

I digitized the 1901 and 1912 editions of industrial inspector reports. There are two main issues which had to be tackled in order to create a panel data set.

First, as industrialization progressed, the number of industrial inspector districts started to increase as well. Unfortunately, this process did not mean the split of pre-existent districts, but new districts were created from parts of different old districts. To overcome this problem, I created a mapping from 1912 to 1901 districts. I exploit the fact that districts were made up of counties and a special survey of all factories in Hungary, which was conducted by the Royal Ministry of Commerce in 1910 and reported the total employment of every industry group-county pair (*A Magyar Szent Korona országai gyáriparának üzemi és munkás-statisztikája az 1910. évről*). Crucially, industrial categorization was the same as in industrial inspector reports or in the census. I assign to every county its share of district-industry group level variables (measured in 1912) determined by county-industry group level employment retrieved from the 1910 survey as weight. Having constructed county-level engine power capacity and employment data by industry groups, I aggregate counties to the 1901 districts as a final step. Note that this imputation might introduce more noise to engine power capacity data than to employment. However, as imputation is done in neighboring counties, I do not expect large differences in the capital-labor ratio within industry groups (Hanson and Slaughter, 2002).

Second, while the 1901 edition of industrial inspector reports is known to encompass all plants subject to 1893/XXVIII (the law regulating the activity of industrial inspectors; Fenyvessy, 1902), principally owing to the proliferation of small engines, not all plants could be visited by inspectors in 1912. In Table A21, it is shown that the share of not inspected plants or employment[53] are not correlated with the instrumental variable.[54] This is important since I only have engine power capacity and detailed employment data (by skill) for inspected plants.

Last, I measure distance to Sáros (/Budapest/Fiume) from each district as a simple county population-weighted (1900) mean of distances from counties which made up a given district. I restrict the sample to industry group-district pairs which had more than twenty employees in 1901 and 1912, had non-zero engine power in 1901, fewer than 25% of plants were closed when visited (1912) and at least half of the factories were visited in 1912, to gain a reliable coverage.[55] Fejér and

[53]By 1912, industrial inspectors had a centrally maintained registry of plants which they were expected to visit. The total number of workers and factories was published together with data on the actually visited ones.

[54]This is true for every weighting used with these data, for instance, total sectoral engine power or employment in 1901. Results not reported.

[55]I drop outlier observations with more than 12,000 HP engine power capacity or more than

Pest-Pilis-Solt-Kiskun counties are dropped because they were in the same district as Budapest in 1901. On the other hand, Sáros county is part of the industrial inspector district around Kassa, so it is included.

In the analyzed sample, (i) 79.7% of all factories; (ii) employing 93.4% of all factory workers were inspected, and (iii) merely 7.8% of factories were closed when visited in 1912 (see Table A21).

I validate the sample in that qualitatively the same results can be estimated from it as from the census on factory employment. In Column 4 of Table A21, the dependent variable is the change in the number of factory workers (as % of district population in 1900). The reduced form coefficient is significant and negative. In comparison with Column 5 of Table 1.3, the coefficient is a magnitude smaller. This is unsurprising since those estimates are not within-industry group. However, the effect of increasing proximity to Sáros by one standard deviation is practically the same with both estimation strategies in terms of standard deviation of the outcome. Moreover, the coefficient increases in magnitude as capital-intensive sectors are omitted. This is in line with the heterogeneity analysis presented in Section 1.6.1.

A.1.3 The Great Hungarian Compass

The Great Hungarian Compass was a yearly publication which provided information on the quasi-universe (a few companies sent their information too late or failed to provide it) of Hungarian public limited companies between the 1870s and the end of WWII. It was published with the support of the Ministry of Commerce which also recommended various authorities to buy it in order to help their administration with its high-quality data. My data set covers all entirely Hungarian and mixed (foreign-Hungarian) registered companies. I include those purely foreign-registered firms (their number is very limited, most firms had mixed registration) which issued their balance sheet separately for Hungary.

Merely the stock of registered capital and year of establishment are available for 1-2 year-old firms. For older ones, their balance sheet is reported. Many mature companies provided data on their dividends, engine power capacity and employment, too. Only the largest firms reported their share price and sales. The entries of companies which reported in the old currency (*forint*s) in 1900 were converted to *crown*s.

The three main elements of book capital are the value of machines, equipment and factory building. Besides these, carriages, fences, fountains, horses, own

7,500 workers in 1901. This results in dropping four outlier observations whose inclusion does not change the results qualitatively but makes the estimation less precise. Setting more restrictive upper bounds to eliminate more potential outliers - such as setting maximum engine power at 7,500 HP or limiting employment at 4,000 workers - does not change the results in any meaningful way.

Table A1: Capital and labor-intensive sectors - 1913 (Great Hungarian Compass)

Industry	Engine power (measured in HP)	Number of workers	Book capital (in 1.000 crowns)	Engine power per worker	Capital per worker
Printing (II/B/a/XII)	562	8 582	10 190	0.07	1.19
Mining (II/A)	3 416	19 457	68 431	0.18	3.52
Lumber (II/B/a/IV)	3 721	12 496	52 141	0.30	4.17
Leather (II/B/a/V)	1 170	2 960	8 150	0.40	2.75
Building materials (II/B/a/III)	10 664	21 863	91 054	0.49	4.16
Textile (II/B/a/VI-VII)	9 978	15 734	54 814	0.63	3.48
Iron and other metals (II/B/a/I)	17 170	15 023	82 923	1.14	5.52
Chemical (II/B/a/X)	10 906	6 205	79 224	1.76	12.77
Paper (II/B/a/VIII)	6 546	3 420	24 739	1.91	7.23
Food and beverage (II/B/a/IX)	40 980	20 929	174 682	1.96	8.35
Machinery (II/B/a/II)	74 469	18 002	123 999	4.14	6.89

Note: all Hungarian firms - i.e. companies in Budapest and Sáros county, too - which reported engine power capacity or employment are included (most of them reported both - non-reporting arises mostly for engine power capacity which might be the consequence of non-mechanized production technology used). If the number of workers was reported as maximum and minimum during a year, I took the average. If engine power capacity was recorded as an interval, the maximum possible performance is used. Capital-intensive sectors are shaded grey.

railway connections to main lines, patents, special rights (mining, forests, etc.), the value of ongoing constructions and investments play a minor role.[56] Houses built for workers are not included. In some cases, machines were grouped together with intermediate inputs - these observations are not part of my book capital measure. The following balance sheet items are considered as equity: registered capital, reserves (for depreciation, dividends, taxes or special reasons), reserves for the benefits of workers (accident, aid or pension funds) and (retained) profit.

In the industrial classification, I mainly follow the classification of book chapters to set industry groups. There are some exceptions though. Wood distilling is reclassified as chemical industry to be in line with the census. Three iron- and steelmakers owning mines (Hernádvölgyi Magyar Vasipari Rt., Gróf Csáky László Prakfalvai Vas- és Acélgyár Rt., Rimamurány-Salgótarjáni Vasmű Rt.) are classified as iron- and steel-making companies (subgroup 20) instead of mining. Last, starch production is treated as part of the chemical and not the food industry, and quarrying and magnesite kilns are classified as group II/B/a/III instead of mining.

There were many large companies which possessed several factories in Hungary.[57] Many of them provided their valuation of each plant on their balance sheet. As default, I use this information as weight to allocate capital which is not specifically assigned to a given plant but only to the company as a whole. Equity and total assets are allocated based on these weights, too. If this information is not available, equal weight is put on each factory. A further empirical concern is that the entry of a public limited company might simply mean the takeover of a former

[56] While calculating equipment and machine capital separately from broadly defined capital would be conceptually superior to proxy for mechanization (Lafortune et al., 2019), I cannot do so because many firms valued their equipment together with the factory building.

[57] I only consider factories - storehouses or shops are excluded because of their marginal value compared to factories.

privately-held firm. Fortunately, the data set provides information on this so that all such cases can be deleted.

Last, a few firms which existed before 1899 and still in 1912 reported their balance sheet exclusively in 1899. Those observations are excluded because their missing balance sheet would imply a false exit in the estimation. Conversely, if a firm existed before 1899 and also in 1912, but only provided its 1912 data, the observation is deleted. These restrictions can be made because the Compass reports all existing public limited companies in a given year, even if they failed to send the required data.

A.1.4 Population decomposition - measurement

Variables used in the population decomposition (Equation 1.4) come from the following sources:

- Population change: county-level population from the 1900 and 1910 censuses
- Births: adding up yearly births between 1901[58] and 1910. Stillbirths are not included.
- Deaths: adding up the number of yearly deaths between 1901 and 1910.
- Net emigration: I use data reported by HRCSO (1918) on Hungarian-born emigrants and return migrants who migrated to/from outside the Austro-Hungarian Empire. For the (internal) Austrian direction, I rely on data provided by the Austrian government on the number of Hungarians residing in Austria from every county (3rd volume of the 1900 census and HRCSO (1918, p. 102-103) in 1910). Migration data on Croatia are from Hungarian censuses as they were published jointly with the core Hungarian data. Note that the immigration/emigration of foreign-born people is not included as that would lead to the double-counting (once as return migrant in the migration statistics and once as foreign citizen in the census) of those who left as Hungarian citizens but returned with foreign citizenship (e.g., young men in order to avoid the draft). As the number of foreigners in Hungary increased only by approx. 22,000 between 1900 and 1910, this omission is most likely not influential.
- Net internal out-migration: I use the county-to-county mobility matrices provided by censuses which are not perfect accounts of internal migration. They record the number of people living in county x who were born in county y. Thus, if a person moved between 1900 and 1910, and died, her move

[58]Recall that the 1900 census uses 31/12/1900 as reference date.

was not recorded, but her death appears in her destination county. The net immigration of non-local born Hungarians is defined as the change in the number of people who were not born in a given county. The measure is net since non-locally born people living in the county in 1900 could move out by 1910. The net out-migration of the local born population is defined as the change in the number of people born in county c who lived in other counties of Hungary.

I compare the actual growth in county population with the values implied by the right hand side of Equation 1.4. While the two measures are highly positively correlated (0.9), two findings are worth mentioning. First, implied growth rates are consistently *larger* than the actual ones for counties bordering Austria, suggesting that data provided by Austrian authorities must have underestimated the true extent of migration to the Austrian part of the Dual Monarchy. Second, implied values are regularly *smaller* than the observed ones in southern/south-eastern counties most affected by emigration to Romania. HRCSO (1918, p. 39*) states that «*it is certain that significantly more people returned from the German Empire and from Romania than indicated by our statistics.*» Consequently, the cause behind the underestimation of population growth most likely lies in the inaccurate recording of the number of return migrants.

A.2 Additional anecdotal evidence and empirical analysis

A.2.1 A literature survey of Hungarian chain migration

Chain migration, the phenomenon that mass migration follows some successful pioneers, and its prerequisites are well-documented in Hungary in the studied time period. Sheridan (1907) documents that emigrants regularly sent home accurate information about the US economic conditions. Tonelli (1908) argues that return migrants and letters sent from the US were the main drivers of mass migration. Neményi (1911) claims that the «snowball effect» was one of the main reasons behind Hungarian emigration. Nagy (1983) stresses the immense role of personal, verbal information exchange at local markets among mostly illiterate people. Puskás (1983) highlights the role of contact with foreign pioneers for northern Hungarians and German-ethnicity citizens. Zahra (2017) writes that letters and remittances from relatives or friends were the most persuasive forms of «propaganda» in East Central Europe. Data recorded at Ellis Island reveal that 80-83% of Hungarian-ethnicity immigrants claimed to come to relatives and 15-18% to join a friend in the period of interest (HRCSO, 1918). Hegedüs (1899), surveying early emigrants, found that 8 out of 10 emigrants followed relatives who were already in the US. However, reducing information frictions was not the only reason which made pioneers influential. They could also finance the journey of their relatives and friends to the US. HRCSO (1918) presents that roughly one-third of all journeys made by Hungarian-ethnicity immigrants were paid by their previously emigrated relatives after 1908. Taken together, various pieces of evidence point towards the crucial importance of chain migration in Hungary.

A.2.2 Further threats to the identification

Growth and emigration in Galicia A potential threat to the identification could be a significantly changing economic environment on the other side of the intra-empire border of Austria-Hungary near Sáros, i.e. in historical western Galicia. However, Galicia was not a major trading partner of Hungary. In fact, the survey of Hungarian international trade between 1888 and 1913 exclusively mentions it in relation to crude oil import and milled rice export. Additionally, Frisnyák (2006) notes that railroad connections to Galicia chiefly served military purposes and were very underutilized. Thus, Galician emigration resulting in a shrinking local market for manufacturing goods and particularly negatively affecting close-by regions around Sáros is unlikely to drive the empirical findings.[59]

Galicia had one booming sector in the period of interest, crude oil extraction (Frank, 2007; Kaps, 2015). It started its dynamic expansion in the second half of the 19$^{\text{th}}$ century. As Galicia was practically the only area where crude oil was found in Austria-Hungary, the boom of this sector did not mean a direct competition to Hungarian firms on the output market. On the contrary, access to oil should have spurred industrial development near Galicia.

Permanent in-migration of poor farmers or agricultural laborers from the impoverished Galicia could also explain the relatively modest agricultural worker losses. However, I do not find a differential change in the number of foreign-born individuals at the county level (not reported), which could otherwise suggest the relevance of this in-migration channel. Additionally, I have not found any sources reporting on workers leaving Hungary to work in the Galician (oil) industry.

A trade shock Another potential confounding effect could be a trade shock spatially differentially affecting capital- and labor-intensive sectors. I show the main traded goods of Hungary in Table A2. These goods constituted 75-80% of Hungarian export and import values prior to WWI. We can observe that the structure of trade barely changed, the share of most categories remained almost constant. Hungary mainly exported raw materials (grains, fat stock, etc.), while imported manufactured goods (textile and leather products, ironware, machines, etc.). Around three-quarters of all exports and imports were traded with the Austrian part of Austria-Hungary as the two parts of the empire constituted a

[59]Recall that there is also no heterogeneous effect by the tradability of sectors.

Table A2: Foreign trade of Hungary - value share of main categories

	Import		Export	
Category of goods	1900	1910	1900	1910
Sugar	1,6%	0,8%	3,5%	3,7%
Grains, legumes, flour	2,0%	4,0%	33,1%	30,2%
Fruits, vegetables	1,9%	2,3%	2,2%	3,5%
Fat stock, mules, horses	2,4%	0,2%	15,4%	18,4%
Animal products	1,8%	1,6%	4,8%	5,1%
Beverages	3,3%	1,4%	3,6%	3,2%
Timber, coal and peat	4,2%	4,7%	7,6%	4,3%
Cotton and cottonwares	15,3%	16,2%	1,1%	2,5%
Flax, hemp and jute raw material/products	3,0%	2,7%	1,0%	1,3%
Wool and woolen products	8,6%	9,9%	1,7%	2,2%
Silk and silk products	4,6%	2,9%	0,9%	0,7%
Clothes	6,1%	6,0%	0,9%	0,6%
Leather and leather products	5,0%	5,8%	1,3%	2,1%
Wood and bone products	2,1%	2,7%	0,8%	1,1%
Iron and iron products	4,7%	5,6%	3,3%	1,9%
Non-precious metals and their products	2,6%	2,7%	0,6%	0,6%
Machines and their parts	4,0%	4,9%	1,6%	1,2%
Musical and scientific instruments	2,1%	2,4%	0,5%	0,5%
Total (million crowns)	*1110,3*	*1852,4*	*1327,4*	*1716,8*

Note: own calculations based on foreign trade data published in the 63rd volume of the Hungarian Statistical Bulletin. Goods are included conditional on surpassing 25.000 crowns in import or export value in 1900.

customs union since the mid-nineteenth century. However, freight costs only marginally changed within the empire since the late 1880s (Figure 2 of Schulze and Wolf, 2012). Outside the empire, increasing trade with Germany characterised Hungarian international trade: imports from Germany increased to 9% of the total import value by 1912. Since a main, direct railway line connected the region of first pioneers to Silesia (the Kassa-Oderberg line), trade with Germany could potentially act as a confounding force. While the four-fold increase in the value of imported coal could stunt Hungarian mining in theory, the access to German coal should foster mechanization in manufacturing. Besides coal, products of the metal industry and machine manufacturing dominated imports from Germany. These could create direct competition to Hungarian capital-intensive sectors but encourage mechanization through imported high-quality machines. Reassuringly, both channels work in the opposite direction to my findings. Furthermore, imports created meaningful competition exclusively to the leather industry across sectors of labor-intensive manufacturing. In conclusion, unobserved trade shocks are unlikely to drive the results.

Border effects Looking at Figure 1.3a, one might conjecture that emigration was positively correlated with proximity to the borders. This could be problematic since border regions are known to grow sluggishly owing to the relatively small market size (e.g., Nagy, forthcoming). However, most borders on the aforementioned figure were actually intra-empire borders with free mobility of capital, goods and people, including the border around Sáros. Only the border to Romania and Serbia meant actual foreign borders (see Figure A2). Moreover, controlling for distance to Budapest, which had a central position in the country, should soak up most of the variation related to borders.

The role of state-owned enterprises Widespread state control in the economy could also affect the interpretation of the results. Nonetheless, the role of state-owned enterprises was rather limited in the period of interest. 27 out of 202 mines with more than twenty employees were state-controlled in 1900.[60] 80 out of 2,049 factories were owned by the state in other sectors of the industry. State ownership was mainly concentrated in gold, salt and silver mining, iron and steel production, and the tobacco industry. Taken together, fewer than one-fifth of all factory workers were employed by state-owned firms in 1900.

A.2.3 Sensitivity of the first stage

I use the state-of-the-art method of Cinelli and Hazlett (2020) to assess the sensitivity of the first stage coefficient. I find that if confounders explained 100% of the residual variance of emigration exposure, they would need to explain at least 47.1% of the residual variance of the IV to fully account for the estimated first stage effect. This sensitivity exercise also shows that unobserved confounders would need to explain at least 43.9% of the residual variance both of the IV and of emigration exposure for the null hypothesis (the true first stage effect is equal to 0) to not be rejected at the significance level of 1%. Consequently, the estimated first stage relationship is fairly stable.

A.2.4 Population growth decomposition

In order to decompose population growth, I use the following accounting identity which describes the main drivers of population change between two time periods:

$$\Delta Population_{ct} = Childbirths_{ct} - Deaths_{ct} - Net\ emigration_{ct} - Net\ internal\ out\text{-}migration_{ct} \quad (1.4)$$

[60]This sample consists of Sáros county as well as Budapest. Source: 5th volume of the 1900 census.

where net emigration is defined as difference between the number of emigrants to foreign countries from county c and the number of return migrants to county c. Net internal out-migration is defined as difference between the number of Hungarian citizens leaving county c for another county and the number of Hungarians moving to county c in a given time period. In what follows, I run the reduced form specification for each of these variables as an outcome. All variables are expressed as % of county population in 1900.

The results of the decomposition are shown in Table A3.[61] First, birth or death rates were not significantly affected by proximity to Sáros. This means that the IV is most likely not correlated with an unobserved negative shock (e.g., famine) which could have led to fewer births, excess deaths, economic decline - and emigration. The imprecisely estimated positive effect on childbirths might have been the result of i) higher wages close to Sáros in most of the period of interest since agricultural earnings positively affect fertility (Ager et al., 2020a); ii) remittances and repatriated savings increasing the welfare of extended families; iii) fewer opportunities in the industry (lower opportunity cost of raising children; Ager et al., 2020a); or iv) a Malthusian response to freed-up resources since the economy just started the transition from a Malthusian economy towards a modern one. Next, net emigration is split into the US and non-US direction (mainly Austria, Germany and Romania). Thus, Column 4 contains the first stage coefficient, while Column 5 shows that the US migratory fever seems to have crowded out other directions. In other words, some potential emigrants, who would have emigrated even in the absence of the US opportunity, simply changed their direction and substituted the US direction for other destination countries, which somewhat dampened the negative effect on population.[62] The fact that the coefficient of net emigration to non-US directions is significant within-region is particularly important. It shows that the effect was a real crowd-out and not simply the negative correlation of proximity to Sáros with south-eastern, Romanian-majority counties (emigration to Romania) or with the western border counties (emigration to Austria).[63]

Last, net internal out-migration has an insignificant coefficient. I analyze the components of this variable in Table A4 in more detail. Columns 1 and 2 show that, if anything, fewer non-local born Hungarians migrated into areas exposed to emigration. Second, reduced out-migration of local-born people seems to have lessened population losses (Column 3), but this finding is not significant

[61] In this section, population growth (1891-1900) is also included as a control variable to increase the precision of point estimates.

[62] This is consistent with anecdotal evidence. «*[In Tolna as] in other counties which sent emigrants to Germany, the spread of emigration to the US had an undoubted effect [i.e. substitution away from Germany to the US as destination].*» Source: HRCSO (1918, p. 112*).

[63] Without region-fixed effects, the coefficient is -.98 (0.23). Thus, the point estimate only mildly changes between the two specifications.

within-region (Column 4). However, out-migration to Budapest was significantly reduced even within-region (Columns 5-6). This is reasonable as Budapest was the center of economic progress these years; thus, open-to-migration people, who received information about opportunities in the US, merely opted for an even more lucrative destination than the capital of Hungary. As Neményi (1911, p. 52) put it straightforwardly: «*[in Hungary] those who are not satisfied in their motherland go to the US instead of a neighboring region.*» In Appendix A.1.4, I discuss measurement issues why Columns 2-6 do not exactly add up to Column 1 in Table A3 ($-2.89 \approx 0.41 - (-0.13) - 4.25 - (-0.61) - 0.26$).

Table A3: Population growth decomposition (1900-1910)

	(1) Population growth=	(2) Births	(3) -Deaths	(4) -Net emigration (US)	(5) -Net emigration (non-US)	(6) -Net internal out-migration
Proximity to Sáros	-2.8915*	0.4119	-0.1330	4.2514***	-0.6086*	0.2594
	(1.5961)	(1.3405)	(0.6827)	(1.3931)	(0.3281)	(1.1946)
Mean of outcome	7.96	38.76	27.02	3.50	0.82	0.99
Standard deviation of outcome	5.27	4.21	2.10	3.44	1.19	3.19
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	62	62	62	62	62	62

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The outcome variable is the change in the number of local population (Column 1; i.e. population growth), the sum of babies born (Column 2) and of deceased individuals (Column 3), the net number of emigrants to the US (Column 4; first stage coefficient) and to other destination countries (Column 5; Austria and Croatia included as well), and the net number of internal out-migrants (Column 6) - all expressed as % of county population in 1900. Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. All specifications control for population growth between 1891 and 1900. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Net internal out-migration decomposition (1900-1910)

	Net immigration of non-local born Hungarians		Net out-migration of the local born		Local born migration to Budapest	
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity to Sáros	-0.3117	-0.4843	-1.0750**	-0.2249	-0.5911***	-0.6648**
	(0.4813)	(0.6074)	(0.5172)	(0.7367)	(0.1489)	(0.2676)
Mean of outcome	2.04	2.04	3.03	3.03	0.71	0.71
Standard deviation of outcome	1.73	1.73	1.93	1.93	0.68	0.68
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	Yes	No	Yes	No	Yes
Sample size	62	62	62	62	62	62

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The outcome variable is the change in the number of Hungarians: who lived in county c but it was not their county of birth (Columns 1-2), who lived outside county c which was their county of birth (Columns 3-4), who lived in Budapest instead of county c which was their county of birth (Columns 5-6). All outcome variables are expressed as % of county population in 1900. Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. All specifications control for population growth between 1891 and 1900. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.2.5 Additional robustness checks

County-level analysis I assess the robustness of the effect on factory employment in Table A5. First, I show that the results are not driven by weighting in Panel A.

Next, one might find the employment threshold for factories (more than twenty employees) arbitrary. Therefore, I collected the number of industrial workers in plants employing 10-20 workers in 1900 and 1910. Column 3 demonstrates that the effect on employment change in these establishments was, if anything, negative as well. Winsorizing the dependent variable at the 5^{th} and 95^{th} percentile barely affects the results. Column 5 shows that the coefficient of interest stays highly significant and even mildly increases when I control for the latitude and longitude of counties, even though the distance-based IV is correlated with latitude (see Figure 1.3a). Though the previous inclusion of region-time fixed effects already suggested it, this empirical exercise corroborates the claim that I do not capture a spurious, North-South growth difference in this analysis. One might also be concerned about the sample size and proper coverage of confidence intervals. I implement wild bootstrap to test if the coefficient of interest is indeed significantly different from zero (Roodman et al., 2018). The calculated p-value (2.3%) shows that the coverage of confidence intervals calculated with the standard technique is slightly too narrow. Another concern about standard errors could be their potential spatial correlation. I calculate the p-value for Moran's I with different decay parameters and functional types. I fail to reject the null hypothesis of no spatial clustering in all cases (exponential and power functional types with 300km and 500km distance thresholds). Therefore, adjusting standard errors following Conley (1999) has merely a second order effect.[64] This is most likely due to the fact that the estimation is implemented in differences rather than in a cross section (Kelly, 2020). I re-do all robustness checks with the inclusion of the lagged outcome as a control variable in Panel B. For instance, the bootstrapped p-value falls well below 1% in this specification. Figure A10 is a non-parametric version of the reduced form regression which demonstrates the absence of strong non-linearity.

I use the share of German-ethnicity population in 1900 as a second instrumental variable next to the baseline one (see Section 1.4.2 for the quasi-exogenous reasons behind early and intensive German emigration). To have a parsimonious specification, I exclusively control for the lagged dependent variable and the initial share of literates (a significant correlate of the share of Germans) in the first columns of Table A6. As expected, the share of Germans predicts more emigration to the US and, reassuringly, a negative effect on factory employment growth. Notice that the implied second stage effect of Germans ($-9\% = -1.27/14.2$) is very similar to that of the baseline instrumental variable ($-9.3\% = -0.51/5.46$), resulting in a second stage estimate in Column 3 which is very similar to the baseline one: a county lost on average nine factory workers following the departure of one hundred

[64] Accounting for potential spatial correlation in the error term with a 300-kilometer threshold, the standard error diminishes from 0.036 to 0.034. If a linear decay of distance is imposed in the spatial correlation structure (Bartlett formula), the standard error further shrinks to 0.032.

Table A5: Robustness checks - county-level analysis

Panel A: preferred specification

	(1) IV(benchmark)	(2) No weighting	(3) SMEs	(4) Winsorization	(5) Latitude/Longitude	(6) S.E. checks
Emigration exposure (US)	-0.0975***	-0.0774**	-0.0041	-0.0985***	-0.1091***	-0.0975***
	(0.0357)	(0.0354)	(0.0027)	(0.0347)	(0.0393)	(0.0357)
Bootstrapped p-value						0.023
Moran's I (p-value)						0.647
Kleibergen-Paap F-statistic	50.40	63.52	50.40	50.40	41.01	50.40

Panel B: lagged outcome included

	(1) IV(benchmark)	(2) No weighting	(3) SMEs	(4) Winsorization	(5) Latitude/Longitude	(6) S.E. checks
Emigration exposure (US)	-0.1237***	-0.1043***	-0.0044*	-0.1230***	-0.1010***	-0.1237***
	(0.0341)	(0.0265)	(0.0026)	(0.0328)	(0.0382)	(0.0341)
Mean of outcome	0.85	0.90	0.11	0.85	0.85	0.85
Standard deviation of outcome	0.82	0.91	0.06	0.77	0.82	0.82
Bootstrapped p-value						0.003
Moran's I (p-value)						0.223
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Kleibergen-Paap F-statistic	49.73	60.50	49.73	49.73	39.34	49.73
Sample size	62	62	62	62	62	62

Note: Unit of observation is county. Robust standard errors in parentheses. Specifications are weighted by county population in 1900 - except for Column 2, which is unweighted. The outcome variable is the change in the number of factory workers (Columns 1-2,4-6) or of workers of establishments with more than 10 but fewer than twenty employees (Column 3) between 1900 and 1910, as % of county population in 1900. The dependent variable is winsorized at the 5th and 95th percentile in Column 4. Bootstrapped p-values are generated by Wild bootstrap (with 999 replications, Rademacher weights and small sample correction). Moran's I is calculated with power functional type (decay parameter=1) and a threshold distance of 500km. Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. The specification in Column 5 includes the latitude and longitude of every county as control. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

emigrants to the US. This conclusion is barely affected by the inclusion of the baseline controls or region-fixed effects (Col. 4-5). Additionally, I find that the overidentifying restrictions are most likely valid because the p-value of Hansen's J-statistic is very high in most specifications.

Another advantage of including the share of Germans is that this variable allows me to estimate the first stage even more precisely (see Figure A9). Therefore, the entire statistical region around Sáros may be omitted to test to what extent counties with the largest emigration exposure drive the results. Notice that the point estimates are not significantly different from the earlier ones, even in restrictive specifications with region-fixed effects included (Col. 6-7).

Table A6: Change in the number of factory workers - two instrumental variables

	First stage	Reduced form	Second stage			Region of Sáros excl.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Proximity to Sáros	5.4559***	-0.5087***					
	(0.6125)	(0.1357)					
Share of Germans (1900; %)	14.2170***	-1.2695**					
	(2.9820)	(0.6291)					
$\frac{\Delta_{1900,1891}\ Factory\ employment}{Population\ (1891)}\%$	-0.3762	0.6931***	0.6564***		0.5956***	0.7546***	0.8500***
	(0.2605)	(0.1309)	(0.1211)		(0.1323)	(0.1488)	(0.1791)
Literates in total population (1900; %)	-0.0371	0.0090	0.0057	0.0128	0.0167***	0.0231***	0.0316***
	(0.0265)	(0.0071)	(0.0063)	(0.0140)	(0.0065)	(0.0070)	(0.0107)
Emigration exposure (US)			-0.0925***	-0.0980***	-0.1021**	-0.1517***	-0.1704***
			(0.0258)	(0.0324)	(0.0420)	(0.0531)	(0.0588)
Mean of outcome	3.50	0.85	0.85	0.85	0.85	0.86	0.86
Standard deviation of outcome	3.44	0.82	0.82	0.82	0.82	0.83	0.83
Baseline controls	No	No	No	Yes	No	No	Yes
Region FE	No	No	No	No	Yes	Yes	Yes
Kleibergen-Paap F-statistic			42.23	43.85	16.84	9.74	12.02
Olea-Pflueger F-statistic			32.64	36.18	13.66	8.47	13.02
Hansen's J-statistic (p-value)			0.92	0.89	0.83	0.46	0.25
Sample size	62	62	62	62	62	55	55

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The outcome variable is the change in the number of factory workers (as % of county population in 1900). Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Counties in the statistical region of Sáros (Right Bank of Tisza) are excluded from the specification in Columns 6 and 7. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Municipality-industry subgroup level sample - nearest neighbor matching

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment effect	-0.179	-0.164	-0.202	-0.141	-0.221	-0.1	-0.225	-0.149
p-value	0.003	0.01	0.004	0.007	0.005	0.24	0.002	0.012
Treatment def.	$< 150\ km$	$< 150\ km$	$< 150\ km$	$< 150\ km$	$< 150\ km$	$< 150\ km$	$< 100\ km$	$7\% <$ emigration exposure
Exact matches (%)	100%	100%	94.9%	99.1%	100%	100%	100%	100%
Sample	Full	Full	5+ locations	Full	Labor-int.	Capital-int.	Full	Full
Treatment type	ATT	ATT	ATT	ATE	ATT	ATT	ATT	ATT
Observations	889	889	780	889	560	329	889	889

Note: The dependent variable is growth of factory employment in a municipality-industry subgroup pair defined as difference in log-employment between 1900 and 1910. Nearest neighbor matching is implemented with the *nnmatch* command in Stata and the aim of finding three close matches for each «treated» observation, using the diagonal matrix of inverse sample standard errors as weighting matrix. Exact matching on industry group and non-exact matching on employment in 1900 are required in every column except for Column 3. Standard errors and p-values are calculated with the *robust(3)* option, allowing for heteroscedasticity, and the estimation is adjusted for the bias caused by non-exact matching. Only observations with fewer than 1.000 workers in 1900 are included. The treatment dummy equals 1 if an observation is located in a county i) which is within 150km from Sáros (Columns 1-6; 177 treated observations), ii) which is within 100km from Sáros (Col. 7; 88 treated observations), iii) which had higher than 7% emigration exposure (Col. 8; 172 treated observations). Column 2 requires non-exact matching on municipality population size in 1900 besides the two other dimensions of matching. Column 3 requires exact matching on industry subgroup instead of industry group. Column 4 estimates average treatment effect instead of average treatment on the treated. The sample is split between labor- and capital-intensive sectors in Columns 5 and 6. Labor-intensive industry groups are II/A and II/B/a/III-VII,XI-XIII. Capital-intensive industry groups are II/B/a/I-II,VIII-X.

Municipality-industry subgroup level analysis I implement nearest neighbor matching using the municipality-industry subgroup panel. The baseline estimate in Column 1 of Table A7 implies that being in an emigration-exposed area reduced employment growth of a municipality-industry subgroup cell by 17.9 percentage points. This effect is comparable to the point estimate in Column 3 of Table 1.9 with a 8% emigration exposure (the mean emigration exposure in the «treated» group of counties was 9-10%). The estimated effect is robust to many specification and treatment definition changes which are presented in Columns 2-8.

Table A8: Change in the supply and price of financial capital

	Total assets of local lenders per capita		Interest rate on savings deposits		Mortgage stock per capita [(6) = (7)+(8)]			
	(1) 1894-1899	(2) 1899-1909	(3) 1894-1899	(4) 1899-1909	(5) 1894-1899	(6) 1899-1909	(7) Land	(8) Housing
Proximity to Sáros	-0.010	-0.007	0.069	-0.076	-0.004	0.003	0.001	0.001
	(0.007)	(0.015)	(0.055)	(0.053)	(0.003)	(0.007)	(0.006)	(0.002)
Mean of outcome	0.02	0.12	0.06	-0.13	0.02	0.06	0.05	0.01
Standard deviation of outcome	0.02	0.06	0.21	0.24	0.01	0.03	0.03	0.01
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	No	No	No	No	No	No
Sample size	62	62	62	62	62	62	62	62

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The outcome variable is change in the total assets of local financial lenders per capita (in thousand crowns, Col. 1-2), in the average interest rate paid on deposits (Col. 3-4), and in the value of local land and housing units acting as mortgage collateral (per capita and in thousand crowns; Col. 5-6 sums the two asset classes, while Col. 7 and 8 decompose the coefficient in Col. 6). Per capita measures are calculated as follows: 1894/1899/1909 values are divided by county population in 1890/1900/1910. All dependent variables are winsorized at the 5th and 95th percentile. Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.2.6 The effect on (the price of) financial capital

Digitizing panel data from a contemporary financial survey (*A Magyar Szent Korona országainak hitelintézetei az 1894–1909. évben*) allows me to analyze the potential effect of emigration on the price and stock of local financial capital.

In the first two columns of Table A8, I show that there is no significant reduced form effect on the change in total assets of lending companies (mainly banks and credit unions, divided by county population) before or after mass migration. Using the total amount of savings deposits and interests due, I calculated the average, county-level interest rate paid by deposit taking institutions. However, I do not find a significant effect on change in the local interest rates (Col. 3 and 4). Even the level of interest rates in 1899 supports the notion of a well-integrated Hungarian financial system: interest rates paid ranged from 4% (5^{th} percentile) to 5.1% (95^{th} percentile) and had a particularly low coefficient of variation (below 10%). Finally, I investigate the effect exerted on the stock of mortgages. This variable is defined as the value of assets (land and housing) in a certain county which was used as collateral to a mortgage. Similarly to the previous estimates, Columns 5-8 do not reveal any significant pattern how emigration might have exerted an impact on local lending. Taken together, this empirical exercise supports the assumption that capital was perfectly elastically supplied in the studied time period.

A.2.7 High- and low-skilled employment

Besides providing information on total employment, industrial inspectors classified the employment of each factory by two categories based on occupation. I label engineers, foremen, machinists and managers as high-skilled (their number was published as a sum), while the rest of employees as low-skilled. Practically none of the industry groups had a higher than 10% employment share of high-skilled workers - their share was between 4-6%.[65] Therefore, defining high- versus low-skilled intensive sectors is futile.

I analyze the effect on high- and low-skilled employment growth in Table A9. The point estimates on low-skilled workers are almost the same as the total employment effects (Table 1.12). This is the consequence of the small number of high-skilled workers. The estimated coefficients show that the growth of high-skilled employment was significantly more reduced than of the low-skilled. This hints at capital-skill complementarity since both the capital-to-low-skilled workers and high-to-low-skilled workers ratio dropped.[66] Capital-skill complementarity is supposed to be associated with the spread of electric motors and modern production methods (Goldin and Katz, 1998; Gray, 2013). In the US, 23% of all engine power capacity in manufacturing was driven by electric motors in 1909 (Goldin and Katz, 1998). Remarkably, this is comparable to Hungarian manufacturing in 1910-12.[67] Supporting the possibility of capital-skill complementarity, Kozári (2009) states that a benefit from the laggard industrialization of Hungary was that firms could apply state-of-the-art, Western production techniques which possibly had high white-collar worker requirements (Chandler, 1977). The lack of earlier built factory buildings fit for harnessing steam power might have helped in electricity and novel production method adoption, too (David, 1990).

Putting these results into a broader context, they align well with the existing literature. The fact that capital was a complementary input to high-skilled workers is consistent with Lafortune et al. (2019). They claim that capital was a stronger complementary input to low-skilled than to high-skilled labor until the 1890s. However, this relationship started to change and capital and low-skilled labor became substitutes. My findings might well capture this shift and we witness the emergence of capital-skill complementarity in Central Europe.

[65] Merely the food and beverage industry had 12-13% high-skilled employees. This is consistent with Katz and Margo (2014a) who show that the share of white-collar workers in the US manufacturing (including artisans, not only factories) did not reach 10% until 1910.

[66] This is only suggestive evidence since I cannot establish that the capital's share in output declined with the high-to-low-skilled workers ratio (Lewis, 2013).

[67] If I subtract the engine power of power plants (assuming zero electric motors used by them) which were included in broad machinery, I find that 28% of the remaining engine power capacity was electricity-driven in 1910 (21% with power plants included; source: the labor survey introduced in Appendix A.1.2). Turning to the industrial inspector reports sample, in which I cannot separate power plants but Budapest is not included, I find a three-fold increase in total engine power capacity but a more than ten-fold jump in electric engine power capacity between 1901 and 1912 (from 3% to 11% of the total). However, the backwardness of Hungary compared to the US manifested itself in the *level* of mechanization as practically all sectors had a significantly lower level of engine power-to-worker ratio (Jerome, 1934).

Table A9: High- and low-skilled employment growth (1901-1912)

	High-skilled employment growth			Low-skilled employment growth		
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity to Sáros	-0.324	-0.311	-0.324	-0.209	-0.194	-0.189
Ln(employment in 1901)		-0.090	-0.010		-0.262	-0.285
Ln(engine power capacity in 1901)			-0.110			0.025
Bootstrapped p-value ($H_1 : \beta_{Prox.\ to\ Saros} \neq 0$)	0.043	0.048	0.057	0.060	0.080	0.078
Bootstrapped p-value ($H_1 : \beta_{high-skilled} < \beta_{low-skilled}$)	0.102	0.098	0.073	0.235	0.103	0.090
Mean of outcome	0.57	0.57	0.57	0.60	0.60	0.60
Standard deviation of outcome	0.46	0.46	0.46	0.43	0.43	0.43
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Industry group FE	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	122	122	122	122	122	122
Number of clusters	15	15	15	15	15	15

Note: Unit of observation is industrial inspector district-industry group. Bootstrapped p-values are generated by Wild cluster bootstrap (with 999 replications, Rademacher weights, small sample correction and clustering at the industrial inspector district (1901) level). Columns 1-3 (4-6) are weighted by high-skilled (low-skilled) employment in 1901. The dependent variable is difference in the log of high-skilled employment (engineers, foremen, machinists and managers; Columns 1-3) and of low-skilled employment (all other workers, Columns 4-6) between 1901 and 1912. The dependent variables are winsorized at the 10th and 90th percentile. Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Labor-intensive industry groups are II/B/a/III-VII,XII. Capital-intensive industry groups are II/B/a/I-II,VIII-X.

A.3 Additional figures

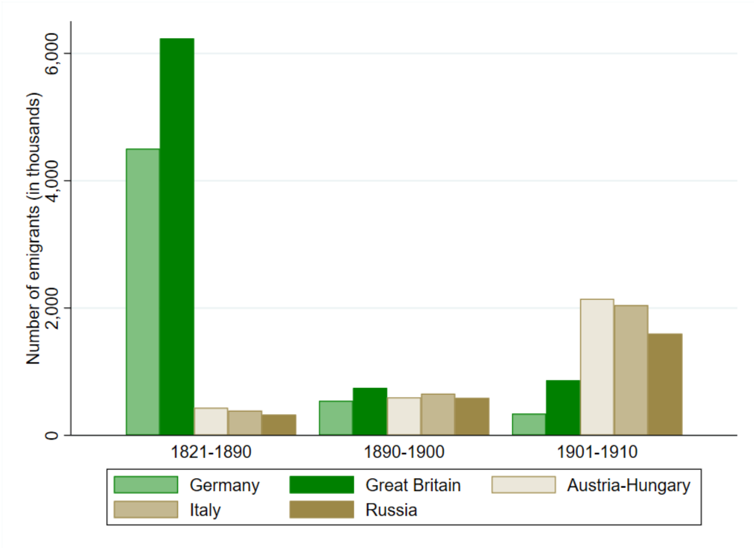


Figure A1: Emigration to the United States by sending countries (1821-1910)

Source: HRCSO (1918).

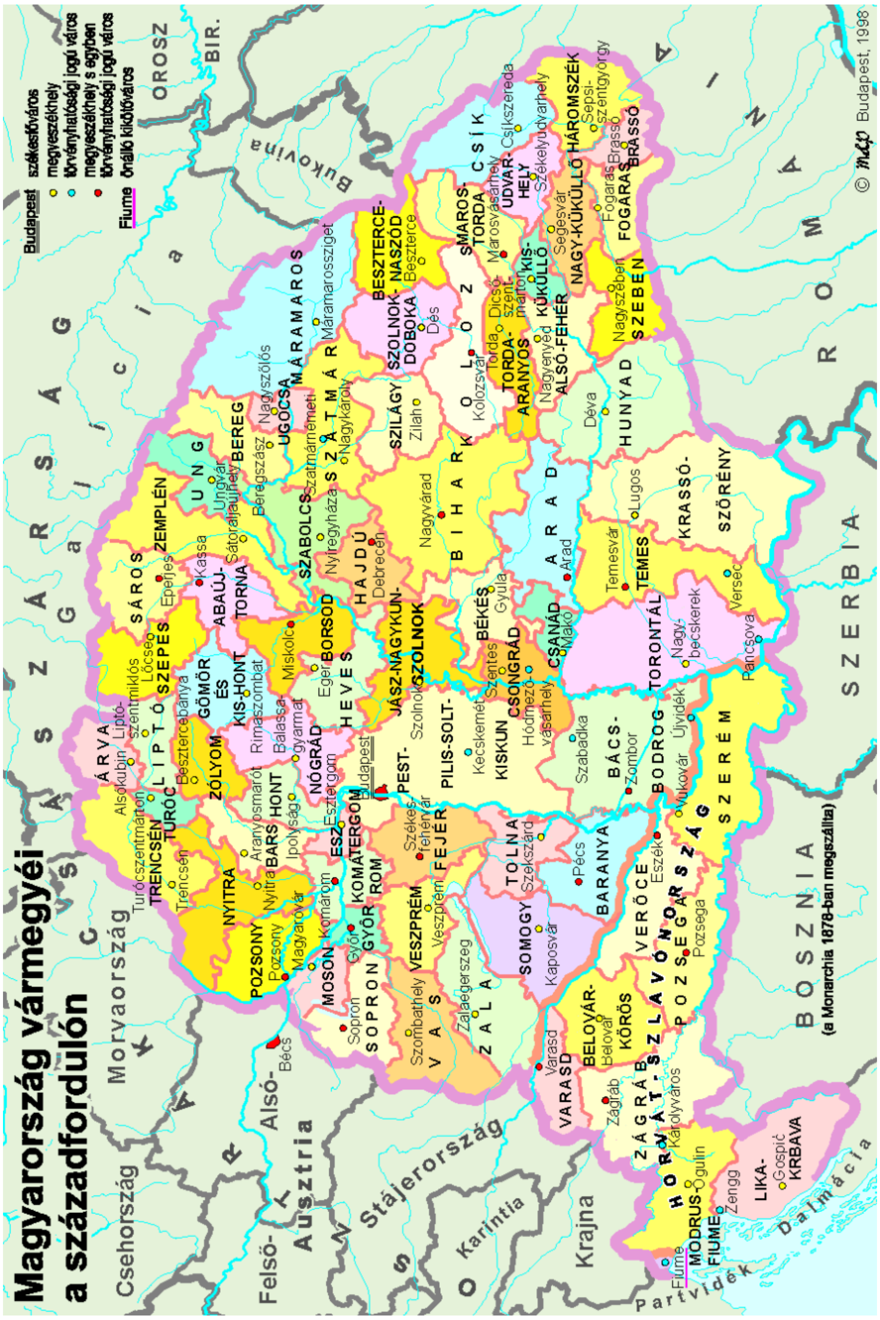


Figure A2: The counties of Hungary before WWI

Source: `http://ishm.elte.hu/hun/maps/1910/varmegy.gif` Accessed 07/07/2020. Counties Belovár-Kőrös, Lika-Krbava, Modrus-Fiume, Pozsega, Szerém, Varasd, Verőce and Zágráb constituted the Kingdom of Croatia-Slavonia which is not part of my estimation sample. Yellow dots indicate county seats, while blue dots indicate royal free cities. Red dot indicates when these two categories coincided.

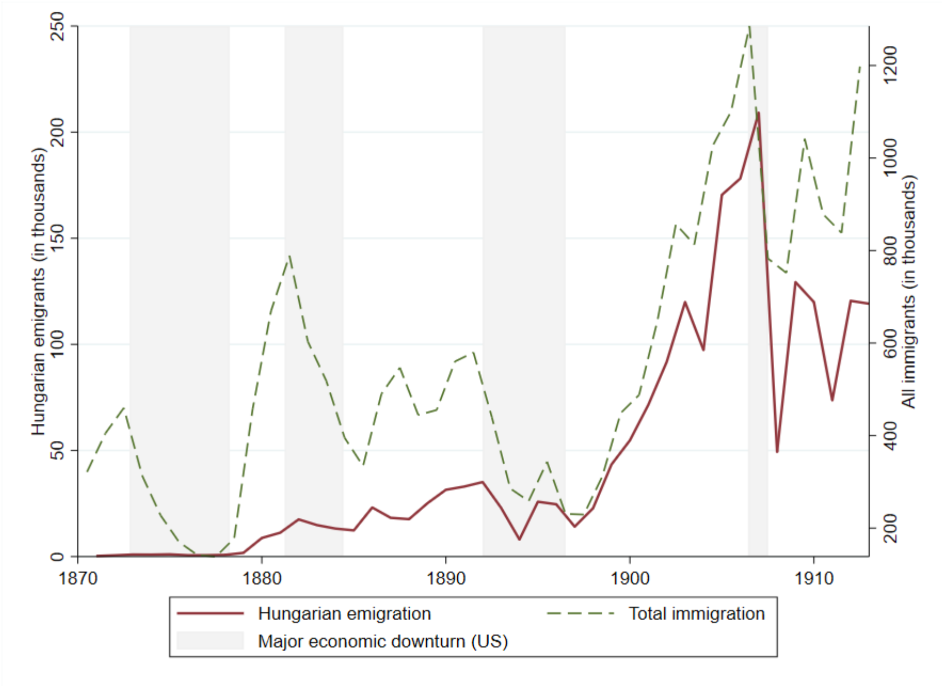


Figure A3: Emigration to the United States (1871-1913)

Note: the source of the number of Hungarian emigrants (left axis) is combined European port data published in HRCSO (1918, p. 150). These data were published in calendar years and also include the citizens of the autonomous Croatia-Slavonia. Data on the total number of immigrants to the US are from the Yearbooks of Immigration Statistics. They were recorded by fiscal year ending on June 30. A recession period qualifies as major downturn if business activity declined by at least 25% as calculated by Zarnowitz (1996). The Panics of 1893 and 1896 are treated as one.

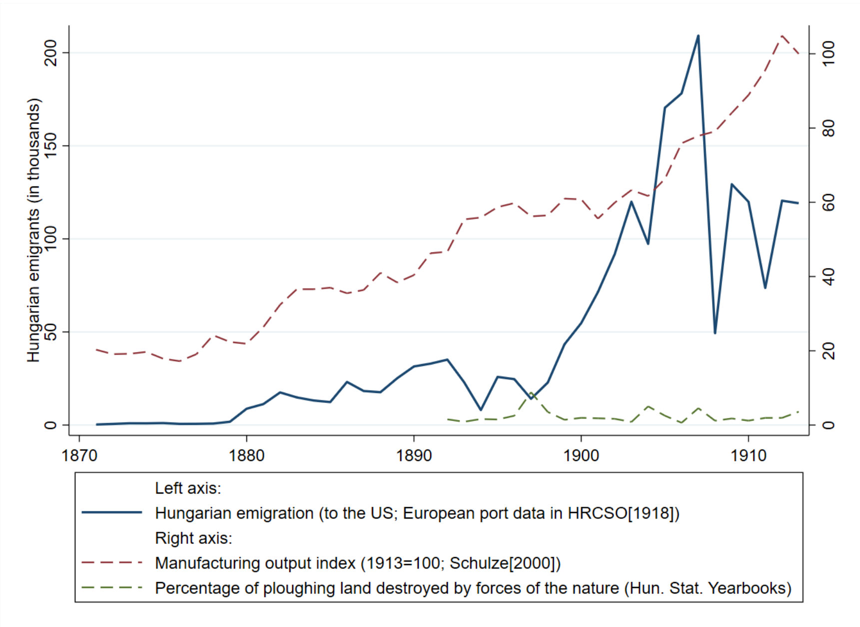


Figure A4: Emigration and the Hungarian economy (1871-1913)

Note: forces of the nature include flood, drought, frost, hail, worms and mice, among others.

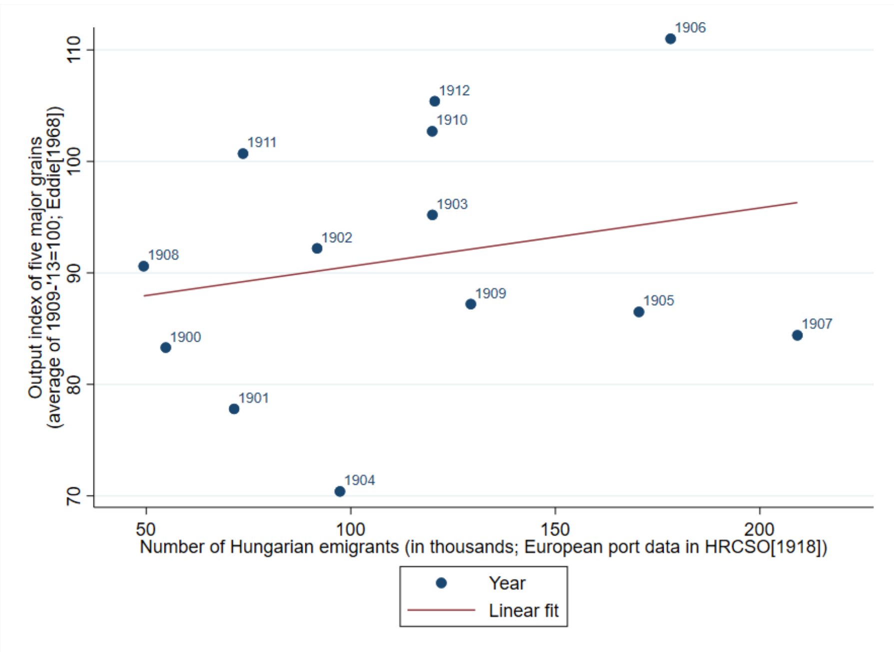


Figure A5: Emigration and agricultural production (1900-1912)

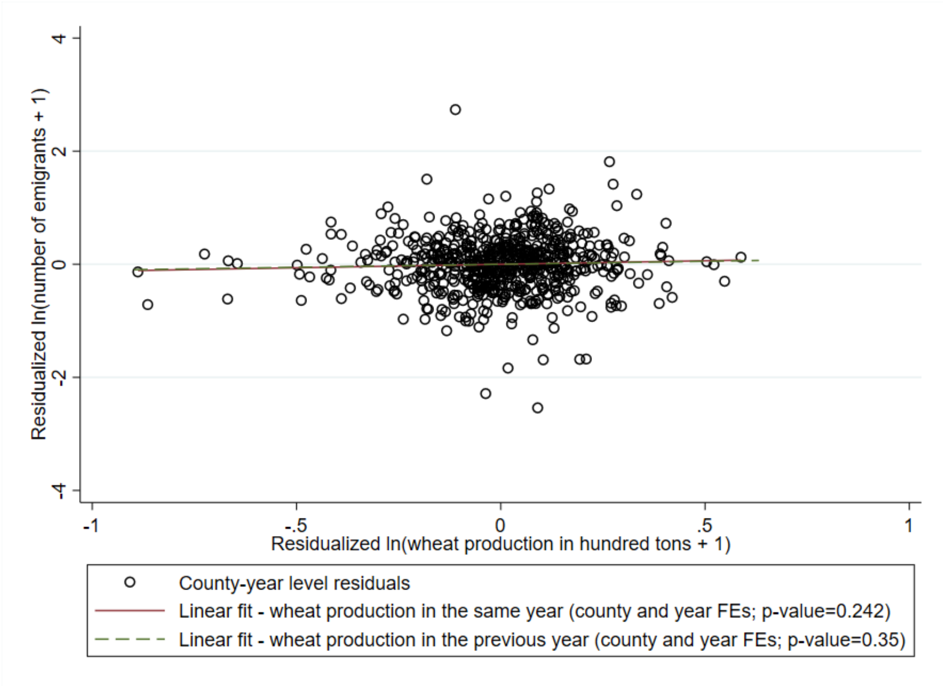


Figure A6: Yearly emigration and wheat production at the county level (1901-1910)

Source: The number of emigrants includes emigrants to the US and to every other international destination (i.e. outside Austria-Hungary). Data on yearly wheat production can be retrieved from statistical yearbooks. The log of yearly emigration and wheat production is residualized using unweighted OLS regressions controlling for county- and year-fixed effects. The three variables - yearly number of emigrants, wheat production in the same and previous year - are winsorized at the 99th percentile. P-values are calculated based on standard errors clustered at the county level. The county-year level residuals (circles) in the figure pertain to the OLS regression between emigration and wheat production in the same year.

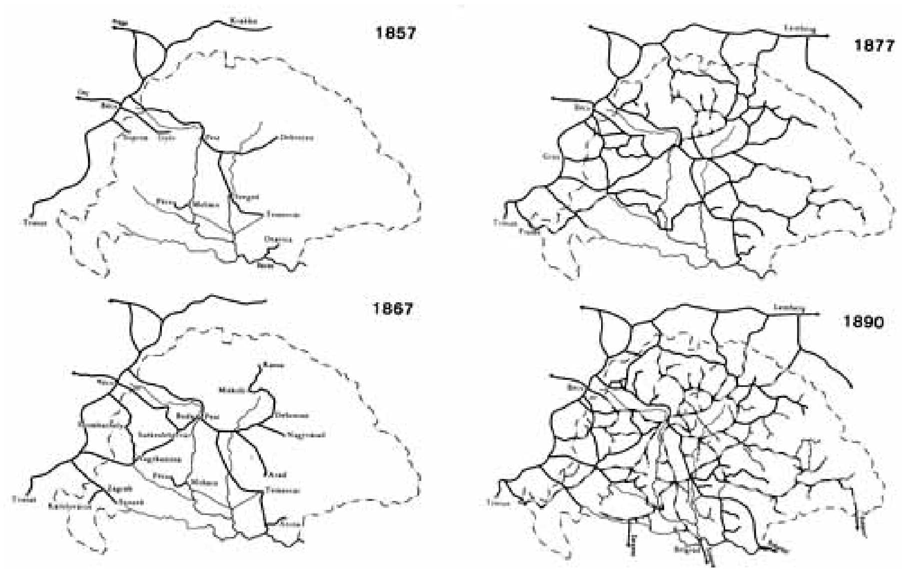


Figure A7: The expansion of Hungarian railway lines

Source: `http://idemonarchia-foldrajz.blogspot.com/2017/10/a-monarchia-vasuthalozata.html` Accessed: 24/02/2020.

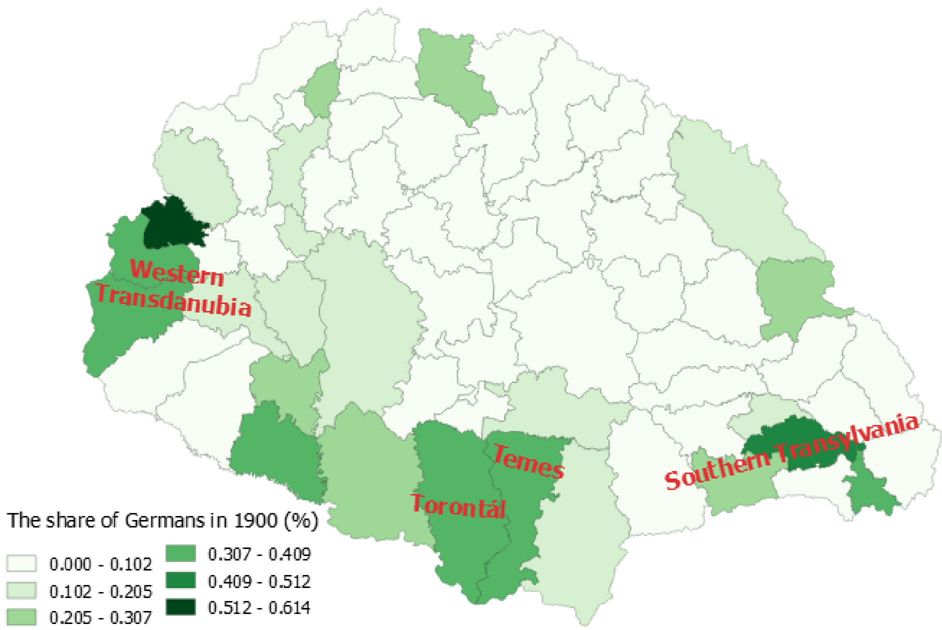
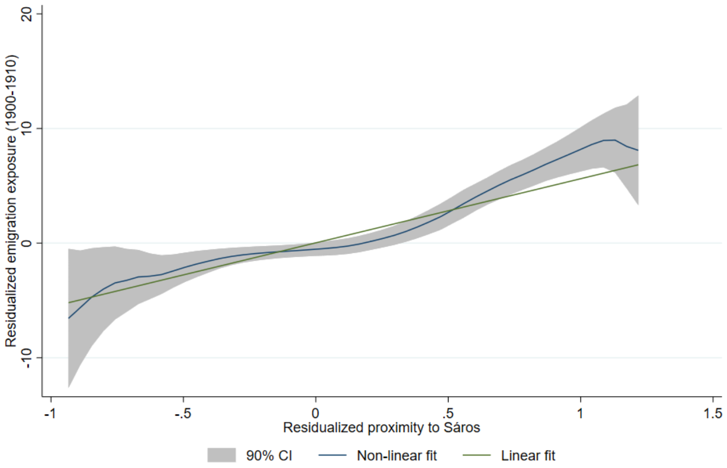
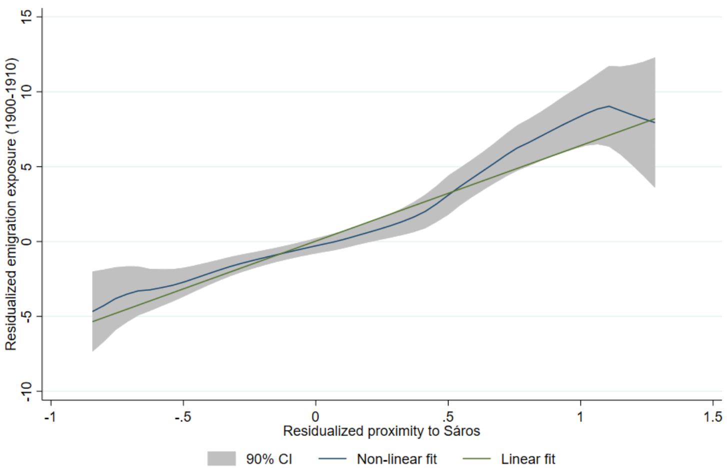


Figure A8: The spatial distribution of German-ethnicity citizens in 1900

Note: data from the census conducted in 1900. Six equal-length intervals.



(a) Preferred specification



(b) Additionally controlling for the share of Germans in 1900

Figure A9: First stage - county level

Note: own calculations. Observations are weighted by county population in 1900. Variables used for the residualization are proximity to Budapest and Fiume/Rijeka and the ratio of industrial workers to all workers in 1891. I use the Stata command *twoway lpolyci* to create 90% confidence intervals with the following options: Epanechnikov kernel function, rule-of-thumb kernel estimator, first degree smoothing. The share of German-ethnicity population in 1900 is included in Figure (b) as an additional control variable. The slope of the simple linear fit is 5.34 and 6.07 in Figure (a) and (b), respectively. Figure (b) is relevant for the analysis in Appendix A.2.5.

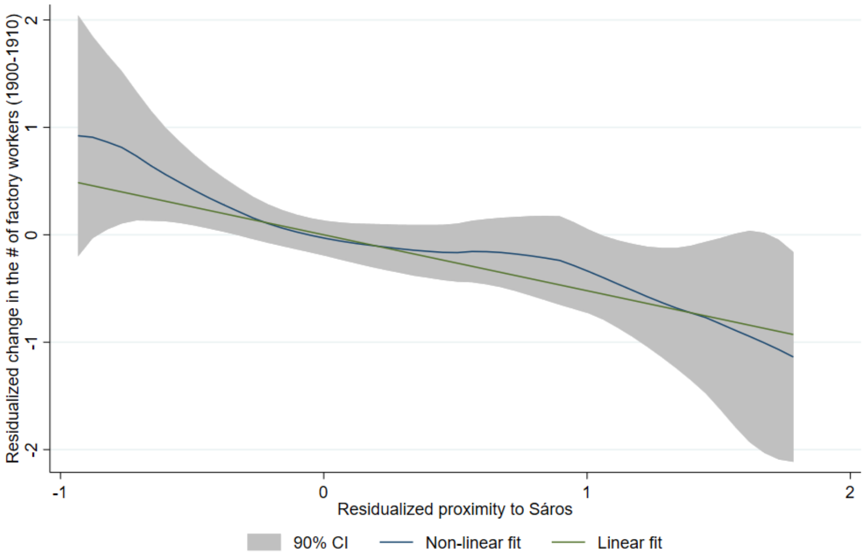


Figure A10: Reduced form - county-level factory employment change

Note: own calculations. Observations are weighted by county population in 1900. Variables used for the residualization are proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. I use the Stata command *twoway lpolyci* to create 90% confidence intervals with the following options: Epanechnikov kernel function, rule-of-thumb kernel estimator, first degree smoothing.

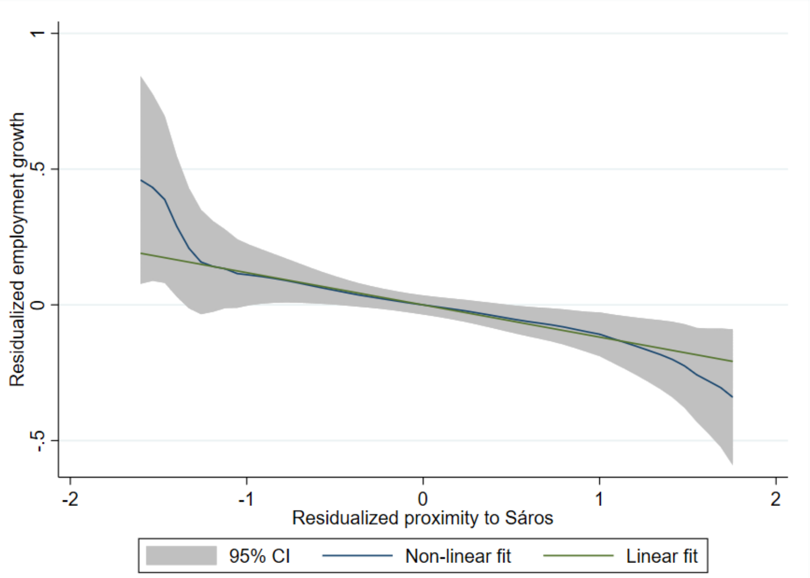


Figure A11: Reduced form - municipality-industry subgroup level factory employment growth

Note: own calculations. Observations are weighted by employment in 1900. Variables used for the residualization are proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. I use the Stata command *twoway lpolyci* to create 90% confidence intervals with the following options: Epanechnikov kernel function, rule-of-thumb kernel estimator, first degree smoothing.

A.4 Additional descriptive statistics

Table A10: Summary statistics - county sample

	Mean	Median	S.D.	Min	Max
Emigration exposure (US)	3.50	2.25	3.44	0.07	15.22
County population (1900; in thousands)	356.02	302.71	207.76	51.92	880.79
Population density (1900)	60.30	61.83	13.67	26.34	85.24
Population growth (1891-1900; %)	9.30	8.27	4.48	0.15	20.20
Literates in total population (1900; %)	50.37	54.56	13.75	17.39	72.13
Change in the number of literates (1891-1900; % of 1891 population)	6.93	6.87	2.15	1.75	15.04
Primary school attendance (1899/1900; %)	81.09	82.80	12.03	42.37	98.29
Financial capital per capita (1899)	0.12	0.11	0.08	0.02	0.71
Share of workers (1900; % of total population)	35.42	35.25	2.21	29.15	43.25
Share of industrial workers (1900; % of all workers)	15.32	14.80	4.08	5.69	28.87
Share of factory workers (1900; % of all workers)	3.47	2.31	3.72	0.12	17.30
Change in the number of factory workers (1891-1900; % of 1891 population)	0.60	0.37	0.77	-0.71	3.81
Share of ploughing land (1895; % of total land)	46.34	45.36	16.55	10.90	83.30
Share of forests (1895; % of total land)	23.42	23.66	15.63	0.56	59.20
Share of infertile land (1895; % of total land)	4.96	4.84	1.63	1.84	13.90
Agricultural mechanization (1895)	0.08	0.08	0.04	0.00	0.20
Percentage of non-Catholic population (1900)	52.06	46.38	29.33	7.83	97.44
Percentage of non-Hungarian ethnicity population (1900)	49.31	46.98	31.35	0.62	98.27
Ethnic diversity (1900; Herfindahl-index)	0.41	0.43	0.21	0.01	0.74
Share of agricultural workers without own plot (1900; % of primary prod. workers)	54.28	56.43	11.30	17.66	75.63
Share of ploughing land with limited salability (1885; % of all ploughing land)	11.23	11.78	5.36	1.93	27.17
Distance to Budapest	218.19	191.91	113.07	39.10	546.86
Distance to Fiume/Rijeka	529.31	531.86	162.97	250.85	889.25
Certified doctors per 10.000 locals (1900)	2.21	2.25	0.62	0.87	4.32
Hospital beds per 10.000 locals (1900)	11.00	8.37	6.87	0.94	39.07
Road density (1900; measured in km per 100 km^2)	16.28	15.55	9.05	2.68	42.43
Railroad density (1900; measured in km per 100 km^2)	5.80	5.90	1.93	1.44	9.98
Telegraph density (1900; number of offices per 1.000 km^2)	10.03	9.85	3.51	3.12	19.72
Observations	62				

Note: all statistics are weighted by county population in 1900. The definition of emigration exposure can be found in the main text. Population density is defined as population per km^2. Literacy means the ability to read and write. Primary school attendance is defined as the share of students who actually attended primary school out of the total number of young people for whom it was mandatory (6-15 y.o.). Financial capital is proxied by the sum of total assets of lending institutions in a county and measured in thousand crowns. This measure is divided by county population in 1900. Agricultural mechanization is defined as the sum of locomobiles, steam-powered plows and threshing machines, and divided by the total ploughing land area measured in iugerums. The H.I. for ethnic diversity is calculated as 1 minus the square of the share of the following ethnic groups: Croatian, German, Hungarian, Romanian, Ruthenian, Serbian, Slovak. Agricultural workers without an own plot are servants and day laborers. Limited salability means church, fee tail (*hitbizomány*), municipal or state ownership. Distance is measured in kilometers from county seats.

Table A11: Largest industries in the municipality-industry subgroup sample

Top 10 - # of municipalities		Top 10 - employment (1900)	
Name of industry subgr.	# of municipalities	Name of industry subgr.	# of workers
Sawmilling	97	Mining (coal)	30.412
Mills (food)	56	Iron and steel production	22.097
Brick production	53	Tobacco production	11.298
Mining (coal)	52	Mining (silver and gold)	10.883
Construction	45	Sawmilling	9.088
Quarrying	34	Sugar production	8.060
Accommodations	28	Mining (iron)	7.729
Iron and steel production	28	Railway workshops	5.627
Mining (iron)	27	Spinning and weaving	4.731
Railway workshops	25	Mills (food)	4.166

Note: own calculations.

Table A12: Largest expansions and contractions - full sample and municipality-industry subgroup panel

Municipality-industry subgr. panel		**Full sample**	
Industry subgr. name	*Employment change (1900-1910)*	*Industry subgr. name*	Employment change (1900-1910)
Top 5 expansions			
Mining (coal)	10,455	Mining (coal)	15,233
Sawmilling	5,911	Sawmilling	14,866
Spinning and weaving	4,479	Construction	7,379
Machine, furnace and ship prod.	3,901	Brick production	6,487
Iron and steel production	3,771	Machine, furnace and ship prod.	6,424
Top 5 contractions			
Mining (gold and silver)	-1,894	Mining (gold and silver)	-1,633
Wagon production	-416	Water regulation	-1,070
Smelting (other metal)	-375	Mining (non-precious metal)	-470
Lime, magnesite and gypsum kilns	-192	Mining (other metal)	-223
Walking stick production	-135	Cattle fattening	-214

Note: own calculations.

Table A13: Dispersion of retail prices and inflation rates - regional markets (1900-1910)

Product name	CoV in 1900	Inflation index (Kassa)	Avg. inflation (excl. Kassa)	S.E. of inflation
Easily tradable				
Bacon (smoked)	0.14	1.54	1.58	0.12
Lard	0.11	1.9	1.6	0.09
Oat	0.10	1.33	1.4	0.08
Rice	0.11	0.96	1.02	0.19
Rye	0.08	1.29	1.35	0.11
Rye flour	0.11	1.17	1.34	0.10
Sugar	0.03	1.01	1	0.05
Wheat flour (low-quality)	0.13	1.74	1.54	0.11
Wheat (low-quality)	0.05	1.54	1.56	0.08
Wine (new)	0.10	0.93	0.97	0.10
Perishable or bulky				
Beans	0.21	1.18	1.51	0.14
Beef	0.22	1.08	1.22	0.11
Bread (wheat)	0.13	1.28	1.33	0.10
Cabbage	0.22	1.04	1.26	0.14
Firewood	0.18	1.27	1.26	0.09
Milk	0.10	1.16	1.3	0.10
Peas	0.34	1.34	1.42	0.15
Pork	0.17	1.64	1.59	0.12
Potato	0.29	1.43	1.3	0.17
Tomato	0.38	0.8	1.08	0.16

Note: CoV: coefficient of variation across regional markets (standard deviation divided by the mean). Inflation indices are calculated as the yearly average market price of 1910 divided by the price of 1900. They reflect medium quality if not noted otherwise. S.E.: standard error. Markets other than Kassa are Arad, Besztercebánya, Brassó, Debrecen, Győr, Kolozsvár, Máramarossziget, Miskolc, Nagykanizsa, Nagyvárad, Pancsova, Pécs, Pozsony, Sopron, Szabadka, Szeged and Temesvár. Budapest is not included.

A.5 Additional regression tables

Table A14: Placebo regressions

Panel A:

	(1) Pop. growth (1869-1900)	(2) Population density	(3) % of county pop. born in the county	(4) Forests	(5) Ploughing land	(6) Agricultural mechanization	(7) Literacy rate	(8) Literacy rate ch. (1891-1900)	(9) Primary school attendance (%)
Proximity to Sáros	-2.288	0.599	-0.530	3.691	-1.713	-0.002	-0.557	0.282	2.279
	(2.435)	(2.337)	(0.913)	(3.170)	(3.451)	(0.009)	(2.255)	(0.633)	(2.143)
Proximity to Fiume/Rijeka	1.227	22.573***	3.168*	-15.440	20.929	0.068**	14.347***	1.188	17.274***
	(5.081)	(6.994)	(1.660)	(9.455)	(12.671)	(0.031)	(4.854)	(1.293)	(5.931)
Proximity to Budapest	8.203**	4.475	-2.962***	-12.791***	10.563**	0.022	8.025***	-1.502***	1.897
	(3.595)	(3.208)	(0.787)	(3.867)	(4.666)	(0.015)	(1.904)	(0.565)	(2.347)
Share of industrial emp. (1891)	-0.642*	-0.022	-0.312**	1.549***	-1.125*	-0.000	1.326***	0.001	1.090**
	(0.351)	(0.534)	(0.140)	(0.579)	(0.635)	(0.002)	(0.403)	(0.089)	(0.457)
Mean of outcome	20.38	60.30	89.37	23.42	46.34	0.08	50.37	6.93	81.09
Standard deviation of outcome	9.10	13.67	3.67	15.63	16.55	0.04	13.75	2.15	12.03

Panel B:

	Financial capital per capita	% of non-Hungarians	Ethnic diversity	Railroad density	Telegraph density	Certified doctors per 10.000 locals	% of non-R. Catholics	Land w/ limited salability (%)	Agric. workers without plots (%)
Proximity to Sáros	-0.026	-7.759	0.004	-0.269	0.507	0.021	-4.950	1.000	3.607
	(0.025)	(7.322)	(0.056)	(0.395)	(0.688)	(0.150)	(4.191)	(0.981)	(2.879)
Proximity to Fiume/Rijeka	-0.030	-12.457	0.173	2.503**	5.777**	-0.343	-55.587***	8.261**	7.640
	(0.050)	(14.354)	(0.127)	(1.045)	(2.263)	(0.370)	(7.656)	(3.114)	(6.557)
Proximity to Budapest	0.005	-32.067***	-0.236***	1.160**	0.602	0.271	-14.218***	2.372	10.111***
	(0.018)	(6.289)	(0.056)	(0.486)	(1.067)	(0.173)	(3.413)	(1.487)	(2.343)
% of industrial emp. (1891)	0.011**	2.933***	0.013	-0.007	0.171	0.080***	-0.764	0.067	-1.193**
	(0.004)	(1.053)	(0.008)	(0.074)	(0.149)	(0.030)	(0.702)	(0.184)	(0.516)
Mean of outcome	0.12	49.31	0.41	5.80	10.03	2.21	52.06	11.23	54.28
Standard deviation of outc.	0.08	31.35	0.21	1.93	3.51	0.62	29.33	5.36	11.30
Sample size	62	62	62	62	62	62	62	62	62

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. Population density is defined as the ratio of county population to area measured in km^2. The outcomes in Columns 4-5 of Panel A can be interpreted as percentage of county area in 1895. Agricultural mechanization is defined as the sum of locomobiles, steam-powered plows and threshing machines divided by the ploughing land area measured in iugerums, recorded in 1895. Literacy means the ability to read and write. Literacy rate is defined as the percentage of literates in county population in 1900. Literacy rate change is the difference in the literacy rate between 1891 and 1900. Primary school attendance is defined as the share of students who actually attended primary school out of the total number of young people for whom it was mandatory (6-15 y.o.), and refers to academic year 1899/1900. Financial capital is proxied by the sum of total assets of lending institutions in the county, measured in thousand crowns in 1899 and divided by county population in 1900. The share of non-Hungarian ethnicity people was recorded in 1900 and can be interpreted as percentage. The Herfindahl-index for ethnic diversity is calculated as 1 minus the square of the share of the following ethnic groups in 1900: Croatian, German, Hungarian, Romanian, Ruthenian, Serbian, Slovak. Railroad (telegraph) density is measured in km (office) per 100 (1.000) km^2. Limited salability means church, fee tail (*hitbizomány*), municipal or state ownership. The area of land with limited salability is measured in 1885. Agricultural workers without an own plot are servants and day laborers, and are defined as percentage of all primary production workers in 1900. Significance levels: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A15: First stage - municipality-industry subgroup level analysis

	(1)	(2)	(3)	(4)	(5)
Proximity to Sáros	4.24***	5.20***	5.34***	5.29***	4.45***
	(0.70)	(0.50)	(0.50)	(0.48)	(0.94)
Proximity to Fiume/Rijeka		3.62***	3.95***	3.46***	1.06
		(1.02)	(1.12)	(1.08)	(1.85)
Proximity to Budapest		-2.67***	-2.70***	-2.61***	-2.03***
		(0.62)	(0.61)	(0.57)	(0.54)
Share of industrial emp. (1891)			-0.05	-0.02	-0.00
			(0.09)	(0.09)	(0.09)
Mean of outcome	3.80	3.80	3.80	3.80	3.80
Standard deviation of outcome	3.90	3.90	3.90	3.90	3.90
Industry group FE	No	No	No	Yes	Yes
Region FE	No	No	No	No	Yes
Sample size	920	920	920	920	920
Number of clusters	60	60	60	60	60
R^2	0.57	0.69	0.69	0.71	0.76

Note: Unit of observation is municipality-industry subgroup. Robust standard errors in parentheses. All specifications are weighted by the number of workers in 1900. The outcome variable is emigration exposure (US) which is defined in the main text. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A16: Disproportionately large industrial worker losses (1900-1910)

	Emigration exposure [US; (1) = (2)+(3)]			Industrial workers	
	(1)	(2)	(3)	(4)	(5)
	Total	Industrial (alternative)	Non-industrial (alt.)		
First stage:					
Proximity to Sáros	5.3386***	0.5713***	4.7673***		
	(0.7520)	(0.1247)	(0.7315)		
Second stage:					
Emigration exposure				-1.6187*	-1.6872**
(US; only industrial workers; alt.)				(0.8471)	(0.6844)
Population growth (1891-1900; %)					0.1806***
					(0.0676)
Mean of outcome	3.50	0.53	2.97	2.31	2.31
Standard deviation of outcome	3.44	0.54	3.03	1.48	1.48
Baseline controls	Yes	Yes	Yes	Yes	Yes
Kleibergen-Paap F-statistic				20.97	20.48
Sample size	62	62	62	62	62
Share of predicted emigrants (%)	100.0	10.7	89.3	.	.

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The outcome variable in Column 1 is emigration exposure, which is decomposed into emigrants who had been employed in the industry (Column 2) and all other emigrants (Column 3). The dependent variable in Columns 4 and 5 is the change in the number of industrial workers (1900-1910), as % of county population in 1900. The dependent variable of Column 2 is referred to as 'Emigration exposure (US; only industrial workers)' in the lower part of this table. Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A17: Factory employment losses and firm dynamics

	(1)+(2)+(3) = (4)			
	(1) Intensive margin	(2) Entry	(3) Exit	(4) Total
Emigration exposure (US)	-0.0478**	-0.0411*	-0.0085	-0.0975***
	(0.0238)	(0.0227)	(0.0079)	(0.0357)
Mean of outcome	0.42	0.57	-0.14	0.85
Standard deviation of outcome	0.57	0.42	0.18	0.82
Baseline controls	Yes	Yes	Yes	Yes
Kleibergen-Paap F-statistic	50.40	50.40	50.40	50.40
Sample size	62	62	62	62

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The outcome variable is the change in the number of factory employees in municipality-industry subgroup cells which existed in 1900 and 1910 (Column 1), in municipality-industry subgroup cells which had non-zero factory employment only in 1910 (Column 2), in municipality-industry subgroup cells which had non-zero factory employment only in 1900 (Column 3). The last column shows the total effect, the sum of the three channels. All outcome variables are expressed as % of county population in 1900. Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A18: Municipality-industry subgroup level employment growth - sectoral split

	All observations			Fewer than 500 workers (1900)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Full sample	Labor intensive	Capital intensive	Full sample	Labor intensive	Capital intensive
Proximity to Sáros	-0.1187**	-0.1495**	-0.0834**	-0.1551***	-0.1831***	-0.1158**
	(0.0447)	(0.0653)	(0.0314)	(0.0409)	(0.0651)	(0.0521)
Mean of outcome	0.15	0.11	0.21	0.26	0.20	0.37
Standard deviation of outcome	0.54	0.61	0.42	0.62	0.67	0.51
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Industry group FE	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	920	575	345	844	534	310
Number of clusters	60	59	56	60	59	54

Note: Unit of observation is municipality-industry subgroup. Robust standard errors, clustered at the county level, are in parentheses. Specifications are weighted by the number of workers in 1900. Columns 4-6 contain observations with fewer than 500 employees in 1900. The outcome variable is difference in the log-factory employment between 1900 and 1910, and it is winsorized at the 1st and 99th percentile. Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Labor-intensive industry groups are II/A and II/B/a/III-VII,XI-XIII. Capital-intensive industry groups are II/B/a/I-II,VIII-X. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A19: Decomposition of factory employment losses by tradability (1900-1910)

	(1) = (2)+(3)			(4) = (5)+(6)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Total	Localized	Dispersed	Total	Localized	Dispersed
Emigration exposure (US)	-0.0947**	-0.0516*	-0.0431**	-0.1206***	-0.0745***	-0.0462**
	(0.0369)	(0.0274)	(0.0216)	(0.0335)	(0.0263)	(0.0205)
$\frac{\Delta_{1900,1891}\ Factory\ employment}{Population\ (1891)}\%$				0.6004***	0.5284***	0.0720
				(0.1382)	(0.1272)	(0.0864)
Mean of outcome	0.80	0.36	0.43	0.80	0.36	0.43
Standard deviation of outcome	0.79	0.60	0.44	0.79	0.60	0.44
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	No	No	No	No
Kleibergen-Paap F-statistic	50.40	50.40	50.40	49.73	49.73	49.73
Sample size	62	62	62	62	62	62

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The outcome variable is the change in the number of factory workers between 1900 and 1910 (as % of county population in 1900). Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Factory employment includes all industry subgroups in which there were factories in at least two municipalities in 1900. Localized (dispersed) sectors are those for which the index of spatial concentration (Ellison and Glaeser, 1997) is above (below) the median index value in 1900. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A20: Agricultural wage growth (1898-1912) - subcounty-level analysis

	(1) 1898-1912	(2) 1898-1912	(3) 1898-1903	(4) 1903-1912
Proximity to Sáros	-0.021	-0.057	0.093***	-0.115***
	(0.030)	(0.054)	(0.016)	(0.029)
Mean of outcome	0.62	0.62	0.02	0.60
Standard deviation of outcome	0.24	0.24	0.19	0.24
Baseline controls	Yes	Yes	Yes	Yes
Region FE	No	Yes	No	No
Sample size	505	505	505	505
Number of clusters	62	62	62	62

Note: Unit of observation is subcounty (*járás*). Standard errors, clustered at the county level, are in parentheses. All specifications are weighted by subcounty population in 1900. The dependent variable is difference in the log-wage of adult male agricultural day laborers in the summer. All outcome variables are winsorized at the 1th and 99th percentile. Baseline controls include proximity to Budapest and Fiume/Rijeka (both measured from subcounty seats), and the ratio of industrial workers to total population in 1900 (measured at the subcounty level). Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A21: Industrial inspector reports - basics

	(1) Factories inspected (%)	(2) Worker coverage (%)	(3) Closed factories (%)	(4) Employment	(5) Employment
Proximity to Sáros	5.500	-1.870	-2.521	-0.051	-0.062
Bootstrapped p-value	0.23	0.35	0.30	0.04	0.01
Mean of outcome	79.72	93.24	7.76	0.11	0.12
Standard deviation of outcome	12.73	8.92	4.89	0.09	0.10
Baseline controls	Yes	Yes	Yes	Yes	Yes
Industry group FE	Yes	Yes	Yes	Yes	Yes
Sample type	All	All	All	All	Labor-intensive
Sample size	122	122	122	122	65
Number of clusters	15	15	15	15	15

Note: Unit of observation is industrial inspector district-industry group. Bootstrapped p-values are generated by Wild cluster bootstrap (with 999 replications, Rademacher weights, small sample correction and clustering at the industrial inspector district (1901) level). Specifications in Columns 1 and 3 are weighted by the number of factories in 1912, Column 2 by employment in 1912, and Columns 4-5 by industrial district population in 1900. The outcome variable is the share of inspected factories (% of all factories; Column 1), share of factory workers inspected (% of all factory workers; Column 2), share of factories closed upon inspection (% of inspected factories; Column 3), and change in the number of factory workers (as % of district population in 1900; Columns 4-5). In Columns 4-5, the outcome variable is winsorized at the 5th and 95th percentile. Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Labor-intensive industry groups are II/B/a/III-VII,XII.

Table A22: Engine power capacity and employment log-levels (1901)

	(1) Log(engine power)	(2) Log(engine power)	(3) Log(employment)	(4) Log(employment)
Proximity to Sáros	0.103	-0.138	0.112	0.063
Bootstrapped p-value	0.53	0.55	0.59	0.76
Mean of outcome	8.18	7.25	7.57	7.47
Standard deviation of outcome	1.08	1.57	0.78	0.86
Baseline controls	Yes	Yes	Yes	Yes
Industry group FE	Yes	Yes	Yes	Yes
Sample size	122	122	122	122
Number of clusters	15	15	15	15

Note: Unit of observation is industrial inspector district-industry group. Bootstrapped p-values are generated by Wild cluster bootstrap (with 999 replications, Rademacher weights, small sample correction and clustering at the industrial inspector district (1901) level). Columns 1 and 3 (2 and 4) are weighted by total sectoral engine power capacity (employment) in 1901. The dependent variable is the log of engine power capacity (measured in horsepower in 1901; Columns 1-2) and factory employment (measured in 1901; Columns 3-4). Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891.

Table A23: Decomposition of capital stock changes at public limited companies (1900-1912)

		Capital stock ch. [(2) $\approx (4) + (5)$]				Other outcomes	
	(1) Book K-to-county population (1900)	(2) Full	(3) Full	(4) Labor-intensive	(5) Capital-int.	(6) Equity labor-int.	(7) Total assets labor-int.
Proximity to Sáros	0.0106	-0.0063	-0.0180	-0.0134**	0.0056	-0.0107*	-0.0212**
	(0.0087)	(0.0065)	(0.0150)	(0.0065)	(0.0061)	(0.0058)	(0.0106)
Book K-to-county population (1900)		0.7937***	0.7889***	0.4834***	0.4136***	0.4279**	0.7344**
		(0.2048)	(0.1899)	(0.1772)	(0.1317)	(0.1653)	(0.3120)
Mean of outcome	0.015	0.032	0.032	0.013	0.020	0.011	0.022
Standard deviation of outcome	0.023	0.031	0.031	0.022	0.022	0.018	0.037
Bootstrapped p-value	0.302	0.350	0.281	0.039	0.419	0.074	0.043
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	No	No	No	No
Kleibergen-Paap F-statistic							
Sample size	62	62	62	62	62	62	62

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The dependent variable is the book capital-to-county population ratio (measured in thousand crowns per capita in 1900; Column 1), change in the capital stock (Columns 2-5), equity (Column 6) and total assets (Column 7) between 1900 and 1912 - all measured in thousand crowns at the county level and divided by county population in 1900. The dependent variables are winsorized at the 5th and 95th percentile. Baseline controls include proximity to Budapest and Fiume/Rijeka, and the ratio of industrial workers to all workers in 1891. Bootstrapped p-values are generated by Wild cluster bootstrap (with 999 replications, Rademacher weights and small sample correction). Labor-intensive industry groups are II/B/a/III-VII,XII-XIII. Capital-intensive industry groups are II/B/a/I-II,VIII-X. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A24: Factory employment losses - additional IV for the level of industrial employment in 1891

	(1)	(2)	(3)	(4)	(5)
Emigration exposure (US)	-0.0975***	-0.0994***	-0.0977***	-0.1086**	-0.1068***
	(0.0357)	(0.0362)	(0.0335)	(0.0545)	(0.0314)
Share of industrial emp. (1891)	0.1020***	0.1098***	0.1090***	0.0804**	
	(0.0302)	(0.0343)	(0.0361)	(0.0331)	
Share of industrial emp. (1900)					0.0894***
					(0.0277)
Mean of outcome	0.85	0.85	0.85	0.85	0.85
Standard deviation of outcome	0.82	0.82	0.82	0.82	0.82
Distance controls	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	No	Yes	No
Sanderson-Windmeijer F-stat. (Emigration exposure)	50.4	41.94	37.04	12.02	29.5
Sanderson-Windmeijer F-stat. (% of industrial emp. in 1891/1900)	-	124.06	89.33	83.19	52.94
Hansen's J-statistic (p-value)			0.90	0.53	0.89
Sample size	62	62	62	62	62

Note: Unit of observation is county. Robust standard errors in parentheses. All specifications are weighted by county population in 1900. The outcome variable is the change in the number of factory workers (as % of county population in 1900). Distance controls include proximity to Budapest and Fiume/Rijeka. The instrumental variables applied are: proximity to Sáros (Columns 1-5), share of industrial employment (1881; Columns 2-5) and share of German-ethnicity population (1900, Columns 3-5). Unlike 1891 or 1900, the share of industrial workers in 1881 is defined as the share of total population, not exclusively of workers. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 2

EMIGRATION AND LOCAL STRUCTURAL CHANGE

THEORY

2.1 Introduction

Mass emigration from Austria-Hungary mainly meant the loss of low-skilled workers and, thus, the scarcity of (potential) manufacturing laborers in affected areas. Therefore, in this chapter, I approach the effect of low-skilled emigration as the effect of a reduction in the local labor endowment[1] and formalize what we learn about its consequences from the empirical facts documented in the previous chapter. In particular, I set out and discuss the conditions under which emigration causes local deindustrialization. Before doing so, recall that a theoretical model should be able to match five main empirical findings concerning the effect of emigration on local industrial growth in the long run: i) disproportionate industrial employment losses; ii-iii) this industrial slowdown predominantly stemming from labor-intensive sectors and initially less industrialized areas; iv) disappearance of early wage gains despite continuing emigration; and v) lack of induced mechanization in emigration-exposed regions.

I interpret these findings through the lens of a small open economy model with external economies of scale in manufacturing. In this model, there are two distinct sectors: agriculture and manufacturing. Agriculture produces with a constant-returns-to-scale production function which has two inputs: labor and land. Manufacturing is assumed to produce using labor and capital and under external

[1] Compared to a simple reduction in the labor force, low-skilled emigration may bring brain gains or remittances. However, these channels imply a positive rather than a negative local effect. See the literature review of the first chapter.

increasing returns. Free labor mobility is assumed across sectors, resulting in a single equilibrium wage. Thus, following a negative labor supply shock, there are two opposing forces in this model. On the one hand, fewer workers lead to lower labor productivity (and wages) in manufacturing as a consequence of scale economies. On the other hand, the decreasing marginal product of labor and fixed land endowment in agriculture imply a wage rise when the number of agricultural workers drops. If the two inputs of agricultural production are poor substitutes, agriculture will lose relatively few workers since the labor input required for a unit of land is stable. Therefore, manufacturing will experience large employment losses and its wage diminishing effect will prevail, discouraging mechanization and pushing some additional workers from manufacturing to agriculture. In an extension to the baseline model, I show that labor-intensive manufacturing shrinks more than capital-intensive manufacturing if the larger exposure of the former to the costs of labor is not «compensated» through stronger scale externalities.

The key assumption of the theoretical interpretation is the presence of external economies of scale in manufacturing. A detailed literature survey in Appendix B.1 shows that positive (industrial) agglomeration effects were already present in many countries in the second part of the 19$^{\text{th}}$ century. In addition, there is a long tradition of assuming an increasing-returns-to-scale production technology to explain differences in the level of development (e.g., Murphy et al., 1989; Matsuyama, 1991; Ciccone, 2002), trade patterns (e.g., Krugman, 1980) or the size of population (Krugman, 1991). I also use the model and my empirical estimates to quantify the strength of the scale externalities by estimating the scale elasticity. Reassuringly, my point estimate (0.1) is comparable to other, more contemporary estimates (0.16 in Bartelme et al., 2019).

Since existing (open economy) models without industrial scale externalities cannot rationalize my findings, the model proposed in this chapter introduces a novel mechanism to the migration literature. Studies of more recent immigration shocks typically find that low-skilled immigration is associated with endogenous production technology changes: increasing labor intensity for a broad set of sectors and at most modest differential impacts by initial sectoral labor intensity (e.g., González and Ortega, 2011; Lewis, 2011; Dustmann and Glitz, 2015; Imbert et al., 2020). These findings can be interpreted in a model featuring capital-skill complementarity (Lewis, 2013), a model of directed technical change (Acemoglu, 1998; Acemoglu and Autor, 2011; Hanlon, 2015) or endogenous choice of technologies with differing labor intensity (Beaudry and Green, 2003, 2005; Caselli and Coleman, 2006). However, my empirical results - lack of induced mechanization and substantial output mix changes - are inconsistent with the predictions of these models. Additionally, my findings cannot be rationalized by other well-known open economy models either. If we think about Hungary as a system of counties trading with each other, canonical Ricardian models (e.g., Dornbusch et al., 1977) imply a higher relative wage and specialization in locally more productive sectors in areas exposed to emigration. The first prediction is empirically rejected and large within-country sectoral productivity differences are unlikely (Hanson and Slaughter, 2002). The Heckscher-Ohlin model implies an expansion in the capital-intensive sector following a reduction in the local labor endowment (Rybczynski effect). However, capital-intensive sectors experienced, if anything, a decline. The finding that industrial sectors relying strongly and weakly on local demand declined to a similar extent suggests that a shrinking home market cannot be the main driving force behind the industrial downturn either (Krugman, 1991).[2]

My findings inform two ongoing debates in low-income countries. First, policy implications derived from more contemporary immigration studies should be carefully applied to current emigration flows from low-income countries as capital-intensive technology adoption might not follow emigration in these settings. Second, the results address the issue of rural-to-urban migration in developing countries. They provide a potential rationale for why, for example, rural Africa is almost devoid of manufacturing as a consequence of strong rural-to-urban migration (Henderson and Turner, 2020).

Related literature By establishing that emigration affected the expansion and composition of the nascent industry in an open economy, this work is related to the broader literature of structural change. The described mechanism is in line with the classical view that industry competes with agriculture for labor (e.g., Matsuyama, 1992). In my model, the industrial sector expands as population grows because its productivity rises, shifting additional labor from agriculture towards industry. This is consistent with the view that the pull force of industrial productivity matters at earlier stages of structural transformation (Alvarez-Cuadrado and Poschke, 2011). Studying the case of a small open economy, Bustos et al. (2016) show that the adoption of a labor-saving technology in Brazilian soy production spurred manufacturing employment by reallocating laborers from agriculture. This effect crucially relies on strong complementarity between labor and land in agriculture which is a defining feature of my model as well. This complementarity plays an important part also in the open economy model of Leukhina and Turnovsky (2016) who find that rapid population growth, which characterized England in the eighteenth and nineteenth centuries, had a pronounced role in raising the manufacturing employment share.

The idea that scale externalities matter dates back at least to Marshall (1890).

[2]As I assume that a Hungarian county was a small open economy (see Section 2.2 for a discussion), closed economy forces causing structural change - differentially growing sectoral productivity which changes relative prices, non-homothetic preferences, etc. - are excluded from the pool of potential mechanisms. Boppart (2014) summarizes these channels. See also Kongsamut et al. (2001), Gollin et al. (2002) and Ngai and Pissarides (2007).

Studies documenting the presence of positive spillovers in industry are surveyed in Rosenthal and Strange (2004) and Combes and Gobillon (2015), and papers studying agglomeration forces in historical settings are discussed in Appendix B.1. My work is closely connected to the literature which quantifies the role of these (industrial) scale economies. Data on sectoral spatial concentration are alone insufficient to identify external scale economies since the observed spatial concentration may be the result of local natural advantages. In recent work, Bartelme et al. (2019) overcome this empirical obstacle using demand-side variation and quantify the strength of sectoral scale externalities for mostly developed countries around 2000. In this chapter, I use supply-side variation and the structure of my model to calculate scale externalities at an earlier stage of development.

The structure of this chapter is the following. Section 2.2 discusses the main theoretical assumptions and their viability. Then, Sections 2.3 and 2.4 present the equilibrium of the model and what happens to this equilibrium following an exogenous negative shock to the labor endowment. Section 2.5 provides three extensions to the baseline model and Section 2.6 quantifies the scale elasticity. Finally, Section 2.7 concludes.

2.2 Main assumptions

2.2.1 Discussion of the main assumptions

Small open economy I think of Hungary as a system of counties which trade with other countries and each other. More specifically, I assume that a Hungarian county was a small open economy where prices were fixed by the rest of the world. This assumption is in line with the absence of large differences in inflation dynamics across regions (see Section 1.6.5) and studies showing a high level of market integration in Austria-Hungary (Good, 1984; Schulze and Wolf, 2012). In particular, Cvrcek (2013) documents that all provinces of the Dual Monarchy, which was a free trade zone as well, shared broadly the same inflation trajectories and the coefficient of variation of the costs of laborers' full consumption basket dropped from 0.25 in 1827 to 0.08 in 1910. The small open economy is also a reasonable assumption since most of the main railway lines were constructed by the 1890s in Austria-Hungary which drastically reduced transportation costs (Schulze and Wolf, 2012). These lines were used for trade intensively, even between municipalities hundreds of kilometres far from each other and especially for within-country shipment of industrial inputs and outputs (in addition to export and import; Frisnyák, 2006).

Perfect labor mobility across sectors and no mobility across counties There is one type of homogeneous (low-skilled) labor in the model which is perfectly mobile across the two sectors, agriculture and manufacturing. This is a common assumption in the literature of structural change (see e.g., Swiecki, 2017), which is also realistic for the case of pre-WWI Hungary (see Section 1.6.2) and results in a single equilibrium wage. For simplicity, I restrict the movement of these workers across counties in the baseline model. This is a reasonable assumption since I study a period when financial intermediation was still rather rudimentary, possibly thwarting internal migration among low-income people with high information frictions (high cost of search) and limited knowledge about distant opportunities.[3,4] Strong preferences to stay with one's own family and friends[5] might have contributed to the low level of realized internal migration as well. I think about the inter-regional labor immobility assumption as a convenient reduced form assumption - locals did not leave for other regions of Hungary because the expected return was too low relative to the costs -, but I show in Section 2.5 that allowing for an imperfectly elastic labor supply may actually amplify the proposed mechanism.[6]

Elastic capital supply and fixed land endowment The other production factor in manufacturing is capital which is assumed to be elastically supplied at a given price, $\overline{r}$. This is a reasonable assumption since 25% of investments were foreign direct investments in the period of interest (Kaposi, 2002, p. 214). The increasing lending activity of banking companies played an immense role in the industrialization of the Hungarian countryside as well (Kaposi, 2002).[7] Corroborating these

[3] See e.g., Chernina et al. (2014), Angelucci (2015), Guriev and Vakulenko (2015), Munshi and Rosenzweig (2016), Bazzi (2017), Gray et al. (2019). Dustmann and Okatenko (2014) discuss the relation between wealth and migration decisions in depth.

[4] Hegedüs (1899, p. 77) writes that «*Hungary is one of the oddest countries...in one county people know less about the neighboring county than about America.*»

[5] See e.g., Diamond, 2016; Knudsen, 2019.

[6] 85-90% of Hungarian-born citizens lived in their *county* of birth in 1910. Not simply permanent, even seasonal internal migration was limited. Only approx. 80,000 workers participated in temporary inter-county migration to get a job during the harvest season in 1900. This figure is dwarfed by the total number of agricultural workers (less than 5%; Bernáth, 1902). Nevertheless, low internal labor mobility was not a unique characteristic of Hungary. Fuchs (2018) studies structural change in Spain and finds a dominant role for spatial mobility costs and limited internal migration during the WWI boom of the country.

[7] While only one-sixth of all outstanding mortgages were for houses or land located in Budapest, financial institutions headquartered in Budapest provided 55% of all mortgages in Hungary in 1909, attesting to the flow of capital from Budapest to the countryside. Furthermore, the price of machinery was probably also relatively inelastic to local factory construction as the equipment was normally imported from Austria or Germany, or produced in one of the main machinery hubs of Hungary (e.g., Arad, Budapest, Győr). For instance, the ratio of imports to Hungarian production was 7:5 for metal- and wood-processing equipment and sewing machines, 1:2 for steam,

facts, I present results which show no significant reduced form effect either on county-level financial capital accumulation or on interest rates in the pre or post period (Appendix A.2.6). Therefore, the reliance of local growth on local capital accumulation must have been limited. As opposed to manufacturing, agriculture is assumed to produce with a local fixed land endowment ($\overline{T}_A$) which has an endogenous price, r_T.

Scale economies in manufacturing I assume that manufacturing exhibits external increasing returns to scale, while agriculture produces with constant returns to scale. Assuming external economies of scale in manufacturing is consistent with studies of (industrial) agglomeration economies in comparable time periods. A detailed literature survey can be found in Appendix B.1.

Additionally, I empirically support the presence of external increasing returns in historical Hungary. Everything else fixed, scale externalities must lead to disproportionately high entry of factories in areas where manufacturing is already present. Column 1 of Table A24 shows that the share of industrial employment in 1891, which is one of the baseline controls, is positively correlated with factory employment growth in 1900-1910. I use the lagged value of this variable measured in 1881 as an instrument in Column 2.[8] The coefficient of interest remains practically unchanged and highly significant. I add the share of German-ethnicity population as a third instrument in the next column which allows me to implement Hansen's overidentification test. Its p-value is above 80% which lends support to the validity of the overidentifying restrictions. Even including region-fixed effects barely changes the results. I replace the 1891 share with its counterpart in 1900 in the last column so that the two decades between the variable of interest and the instrument guarantee that no short- or medium-term shocks to factory employment can explain the results. In sum, the robustly estimated positive effect of the level of industrial development on subsequent factory employment expansion supports the existence of external increasing returns.

2.2.2 Primitives of the economy

The small open economy assumption implies that $p_M = \overline{p}_M, p_A = 1$, where I normalize the price of the agricultural good to be equal to 1.

The manufacturing sector is assumed to produce with the following technology:

$$Y_M = v \cdot F(K, L_M) = v \cdot \left[\alpha K^{\frac{\sigma_M - 1}{\sigma_M}} + (1-\alpha) L_M^{\frac{\sigma_M - 1}{\sigma_M}}\right]^{\frac{\sigma_M}{\sigma_M - 1}} \tag{2.1}$$

gas or combustion engines and locomobiles, 2:5 for steam boilers, and 1:3 for electric engines and dynamos in 1899.

[8]More precisely, I calculate the ratio of industrial workers to county population in 1881.

where $F(K, L_M)$ is a constant elasticity of substitution (σ_M) production function which combines capital (K) and labor (L_M). v denotes scale externalities, which are assumed to be exogenously given when firms maximize their profit. Scale externalities take the following standard functional form: $v = L_M^\gamma$. That is, they are simply a function of the local level of manufacturing employment and act as a Hicks-neutral productivity shifter. The extent to which they boost productivity depends on the scale elasticity, γ.

The other sector, agriculture produces with standard CES technology:

$$Y_A = F(T_A, L_A) = \left[\beta T_A^{\frac{\sigma_A - 1}{\sigma_A}} + (1-\beta) L_A^{\frac{\sigma_A - 1}{\sigma_A}}\right]^{\frac{\sigma_A}{\sigma_A - 1}} \tag{2.2}$$

where $F(T_A, L_A)$ exhibits constant returns to scale and has two inputs, land (T_A) and labor (L_A).

2.3 Equilibrium

In what follows, I assume that there is a single stable equilibrium in which both sectors employ a positive number of workers. The related discussion of equilibrium stability and uniqueness is presented in Appendix B.2.

Turning to the equilibrium on different markets, perfect competition on the market of agricultural and manufacturing output implies that the marginal cost of both goods must be equal to their price. Since scale economies are external in manufacturing, the equilibrium condition can be written in terms of the unit cost function (c_i) and price in both sectors:

$$c_M(w, \overline{r}, v) = \frac{c(w, \overline{r})}{v} = \overline{p}_M \tag{2.3}$$

$$c_A(w, r_T) = \overline{p}_A = 1. \tag{2.4}$$

The market clearing conditions are the following:

$$\overline{L} = L_M + L_A = a_{ML} \cdot Y_M + a_{AL} \cdot Y_A \tag{2.5}$$

$$\overline{T} = a_{AT} \cdot Y_A \tag{2.6}$$

$$K = a_{MK} \cdot Y_M \tag{2.7}$$

where a_{ij} denotes unit factor demand in sector i for input j. In this economy, the equilibrium is defined in the following way:

Definition 1. *Given production functions in Equations 2.1 and 2.2, the two-sector economy is in equilibrium if:*

- *given world prices for each sector* $\{\overline{p}_A; \overline{p}_M\}$*, the representative perfectly competitive firm in each sector maximizes its profit,*
- *capital, labor and land markets clear.*

2.4 Emigration in the model

Next, I turn to deriving the predictions of the model about the impact of emigration (an exogenous change in the local labor endowment, $\overline{L}$) using Jones's (1965) algebra. First, I derive the effect of population change on the wage because the wage change is directly related to changes in sectoral inputs (see below).

Proposition 1. *The effect of a change in the local labor endowment on the wage can be expressed as the weighted harmonic mean of sectoral inverse labor demand elasticities with the weights being sectoral employment shares:*

$$\hat{w} = \frac{1}{\lambda_{ML} \cdot \frac{\theta_{ML}}{\gamma} - (1-\lambda_{ML}) \cdot \frac{\sigma_A}{\theta_{AT}}} \cdot \hat{\overline{L}} = \frac{1}{\frac{\lambda_{ML}}{\frac{\partial lnw}{\partial lnL_M}} + \frac{1-\lambda_{ML}}{\frac{\partial lnw}{\partial lnL_A}}} \cdot \hat{\overline{L}},$$

Proof. See Appendix B.5. □

where λ_{ML} represents the initial share of employment in manufacturing, θ_{ij} is the initial cost share of input j in sector i, and $\hat{x} = \frac{dx}{x}$. This proposition highlights that there are two opposing forces determining the ultimate wage effect. On the one hand, a diminishing employment in agriculture leads to higher wages due to the immobile land endowment and constant returns to scale, which imply a decreasing marginal product of labor in this sector. On the other hand, a drop in manufacturing employment reduces the productivity of manufacturing laborers through scale economies, resulting in lower wages. The next proposition lays out the condition under which the latter, manufacturing effect prevails and describes the consequences of this manufacturing downturn.

Proposition 2. *If labor and land are sufficiently strong complements in agriculture, i.e.*

$$\frac{\lambda_{ML} \cdot \theta_{ML} \cdot \theta_{AT}}{(1-\lambda_{ML}) \cdot \gamma} > \sigma_A,$$

then:

i) manufacturing contracts both in terms of capital and employment;

ii) agricultural employment increases;

iii) the equilibrium wage drops;

iv) the capital-labor ratio in manufacturing declines:

$\frac{\hat{K}}{\hat{L}_M} = \frac{(\sigma_M + \frac{\theta_{ML}}{\gamma}) \cdot \hat{w}}{\frac{\theta_{ML}}{\gamma} \cdot \hat{w}} = \frac{\sigma_M \cdot \gamma + \theta_{ML}}{\theta_{ML}}$;

v) the magnitude of the (wage) effect is decreasing in the initial employment share of manufacturing (λ_{ML}):

$\frac{\partial^2 \hat{w}}{\partial \hat{\bar{L}} \; \partial \lambda_{ML}} < 0.$

Proof. See Appendix B.5 for the proof of i)-iv). For v), take the second partial derivative of the expression in Proposition 1 with respect to λ_{ML}. □

If the local labor endowment decreases and labor and land are poor substitutes,[9] agriculture will lose relatively few workers since the labor input required for the fixed land endowment is rather stable. Therefore, manufacturing will experience large employment losses. As capital and labor are q-complements in manufacturing, the employment decline will trigger a reduction in the capital stock as well. Additionally, as manufacturing exhibits scale externalities, total factor productivity and, thus, the wage will decline.[10] This wage drop in manufacturing will attract some workers to agriculture which will end up with more workers compared to the initial situation. Moreover, the wage decline makes labor relatively cheaper than the fixed-price capital. Consequently, manufacturing firms substitute some capital for labor in the long run, decreasing the capital-labor ratio. In addition, initially less industrialized places (low λ_{ML}) are predicted to experience a more pronounced negative effect. This happens because marginal employment changes cause more modest *proportional* productivity and, therefore, wage effects in manufacturing when initially many workers are employed in this sector. In conclusion, the model is able to rationalize all main empirical findings.

Figure 2.1 provides additional graphical intuition to grasp how the equilibrium changes in the aftermath of emigration. The wage in agriculture behaves as usual and declines in the number of agricultural workers. However, the wage offered by manufacturing firms also decreases in the number of agricultural workers because more agricultural laborers imply less workers in manufacturing and, therefore, weaker scale economies. As a consequence of emigration, the manufacturing wage curve experiences a leftward shift. This relocates the economy into a new

[9]Leukhina and Turnovsky (2016) set the elasticity of substitution in agriculture equal to 0.25 in pre-1920 England. Bustos et al. (2016) assume that the agricultural elasticity of substitution is smaller than the land share of agriculture. The latter is known to be around 0.3-0.4 (Weil and Wilde, 2009). Even under conservative parameter assumptions, Proposition 2 holds if σ_A is below 0.5. Reassuringly, the estimated elasticities are smaller than 0.5 in the literature (Binswanger, 1974; Salhofer, 2001).

[10]In the absence of scale externalities ($\gamma = 0$), the capital-labor ratio and wage would remain constant owing to $\bar{r}$.

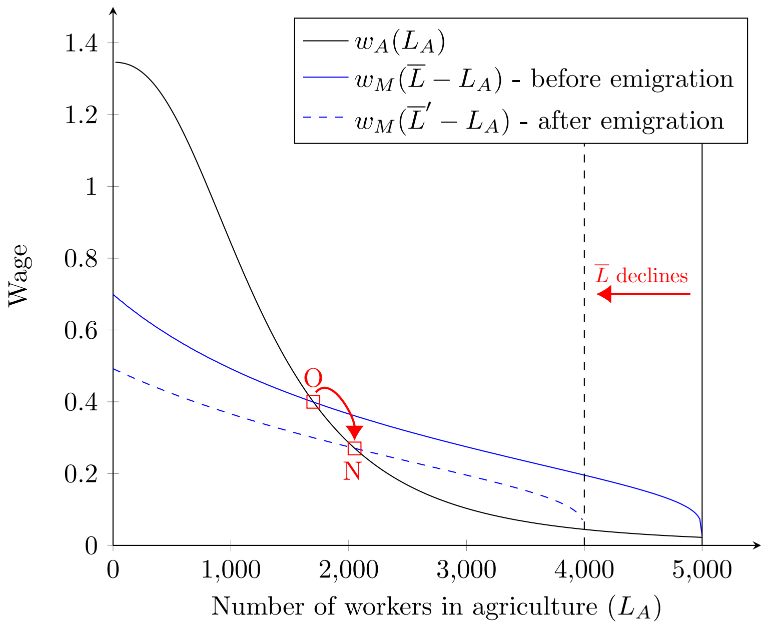


Figure 2.1: The effect of emigration

Note: $\overline{L}$=5,000 and $\overline{L}'$=4,000. The wage equals the marginal product of labor in each sector. Red squares denote the internal equilibrium before and after emigration. O and N denote the old and new equilibrium.

equilibrium with a lower wage and more agricultural workers (from O to N). Note that emigration may be large enough to make the preindustrial, agriculture-only equilibrium the only stable equilibrium. This could happen when the two wage curves do not intersect anymore due to a substantial leftward shift of w_M. However, this hypothetical scenario is rejected by the data.

2.5 Extensions

Capital- and labor-intensive manufacturing The baseline model cannot rationalize industrial output mix changes because it comprises a single manufacturing sector. However, if we assume two manufacturing sectors (the second being labor-intensive) and that scale externalities do not spill over from one sector to the another one, the following strong complementarity condition can be derived:

Proposition 3. $\frac{\partial w}{\partial \overline{L}} > 0 \iff \frac{\lambda_{ML_1} \cdot \theta_{ML_1} \cdot \theta_{AT}}{\lambda_{AL} \cdot \gamma_1} + \frac{\lambda_{ML_2} \cdot \theta_{ML_2} \cdot \theta_{AT}}{\lambda_{AL} \cdot \gamma_2} > \sigma_A$.

Proof. See Appendix B.6. $\square$

The intuition behind this proposition is exactly the same as in the case of a single manufacturing sector (Proposition 2). The more insightful proposition is related to the condition under which labor-intensive manufacturing declines more than the capital-intensive one.

Proposition 4. *Labor-intensive manufacturing declines more in terms of employment if and only if scale externalities are not substantially larger in labor-intensive manufacturing:*
$\frac{\hat{L}_{M_2}}{\hat{L}_{M_1}} > 1 \Longleftrightarrow \frac{\theta_{ML_2}}{\theta_{ML_1}} > \frac{\gamma_2}{\gamma_1}$.

Proof. Using the previously derived expression $\hat{L}_M = \frac{\theta_{ML}}{\gamma} \cdot \hat{w}$ for both sectors of manufacturing. □

This condition entails that labor-intensive manufacturing declines more if its higher exposure to labor is not «overcompensated» with productivity gains from scale externalities. Turning to its plausibility, we have $\frac{\theta_{ML_2}}{\theta_{ML_1}} > 1$ by the assumption that the second sector is labor-intensive. External increasing returns can exacerbate or dampen this difference in labor intensity. However, capital-intensive sectors (steel-making, chemical industry, machinery, etc.) are known to be the transformative sectors of the Second Industrial Revolution (Chandler, 1994; Mokyr, 1999a), which suggests stronger positive externalities for them. Empirically, Henderson et al. (2001) or Bartelme et al. (2019) find stronger external scale economies for capital-intensive manufacturing. Considering these findings, $\gamma_1 > \gamma_2$, i.e. the capital-intensive sector is characterized by stronger externalities, is a tenable assumption which guarantees that Proposition 4 holds. Therefore, a small drop may occur in the capital-intensive sector, while the labor-intensive one experiences a considerable decrease.

Imperfectly elastically supplied labor The model easily lends itself to an extension incorporating imperfectly elastically supplied labor in the following log-additive form:

$$lnL = ln(\phi\left(\frac{w}{\overline{w}}\right)^{\epsilon}) = ln\phi + \epsilon \cdot \ln\left(\frac{w}{\overline{w}}\right),$$

where $\overline{w}$ can be interpreted as a reference wage (e.g., wage in Budapest) and ϕ as the total number of workers when the local wage is equal to the reference wage. Emigration can be interpreted as a shock to ϕ under these assumptions. This leads to a modified relationship between the wage and labor endowment:

$$\hat{w} = \frac{1}{\lambda_{ML} \cdot \frac{\theta_{ML}}{\gamma} - (1 - \lambda_{ML}) \cdot \frac{\sigma_A}{\theta_{AT}} - \epsilon} \cdot \hat{\phi}.$$

The only difference compared to the baseline case (Proposition 1) is the appearance of the labor supply elasticity in the denominator (originally $\epsilon = 0$, as labor is immobile inter-regionally) which makes the strong complementarity condition more demanding:

$$\frac{\lambda_{ML} \cdot \theta_{ML} \cdot \theta_{AT}}{(1-\lambda_{ML}) \cdot \gamma} - \epsilon \cdot \frac{\theta_{AT}}{(1-\lambda_{ML})} > \sigma_A.$$

In the baseline model, emigration reduces the local labor endowment and the wage declines in the long run. In this extension, some additional local workers will then leave the region in search of higher wages, potentially for the United States. This effect is absent in the case of inter-regional labor immobility and exacerbates the negative effect of emigration provided the strong complementarity condition holds.

Agricultural machinery While the baseline model does not allow for capital use in agriculture, agricultural machinery (e.g., harvesters, threshing machines, tractors, etc.) might be purchased to substitute for low-skilled labor, especially in modern times. Therefore, I derive the model with a nested CES production function in agriculture, including elastically supplied agricultural machinery as a third input. The new production function takes the following form:

$$Y_A = F(T_A; (K_A, L_A)) = \left[T_A^{\frac{\sigma_1 - 1}{\sigma_1}} + (K_A^{\frac{\sigma_2 - 1}{\sigma_2}} + L_A^{\frac{\sigma_2 - 1}{\sigma_2}})^{\frac{(\sigma_1 - 1)\cdot\sigma_2}{\sigma_1\cdot(\sigma_2 - 1)}} \right]^{\frac{\sigma_1}{\sigma_1 - 1}} =$$

$$\left[T_A^{\frac{\sigma_1 - 1}{\sigma_1}} + X^{\frac{(\sigma_1 - 1)}{\sigma_1}} \right]^{\frac{\sigma_1}{\sigma_1 - 1}},$$

where K_A represents agricultural machinery, and σ_1 and σ_2 stand for the elasticity of substitution in the outer and inner nest, respectively. Under this new assumption, the relationship between wage and labor endowment changes takes the following form:

Proposition 5. $\hat{w} = \frac{1}{\lambda_{ML} \cdot \frac{\theta_{ML}}{\gamma} - (1-\lambda_{ML}) \cdot \left[\delta_K \cdot \sigma_2 + \sigma_1 \cdot \left(\delta_w + \frac{\theta_{AL}}{\theta_{AT}} \right) \right]} \cdot \hat{\bar{L}}.$

Proof. See Appendix B.7. □

It is easy to see that, when $\delta_K = 0$ (share of agricultural machinery in the production of the inner nest's output) and $\delta_w = 1$ (share of labor costs in the total costs of the inner nest), we are back to the baseline model without agricultural capital. The new theoretical insight is that, all others being equal, the likelihood that the strong complementarity condition holds (in this case, the right-hand-side coefficient must be positive, i.e. $\partial\hat{w}/\partial\hat{\bar{L}} > 0$) declines in the initial level of agricultural mechanization (δ_K) and the elasticity of substitution between labor and machinery in agriculture (σ_2). The interpretation of this result is that agricultural machines essentially break the strong link between land and the number of workers required to cultivate it. In fact, equilibria in which both sectors employ workers and strong complementarity is *not* satisfied are unstable.[11] Thus, starting from an unstable equilibrium, the economy may end up in a manufacturing-only equilibrium at the end of a dynamic process since increasing wages implied by scale economies can attract all workers to industry. The small open economy assumption plays a crucial role in this outcome because it does not let the relative price of the agricultural good increase as agriculture shrinks.

In the late Austro-Hungarian era, the main labor-saving technology was the steam-driven threshing machine, whose spread was well underway even before mass migration started (tractors became widespread only after WWI). Consequently, both the initial level of agricultural mechanization and the availability of machines substituting for laborers were most likely limited (Kaposi, 2002), resulting in a low δ_K and σ_2.[12]

2.6 Quantifying the scale externalities

The relationship between manufacturing capital and labor in Proposition 2 allows me to estimate the scale elasticity using the point estimates on manufacturing employment and engine power capacity growth in Table 1.12 (Columns 1 and 3). Therefore, I only need an assumption on the elasticity of substitution and on labor's share in manufacturing, respectively. The elasticity of substitution for my preferred estimate ($\sigma_M = 2.44$) is the average value of the elasticity of substitution estimated between unskilled labor and capital for the United States in 1860-1880 and 1890-1930 by Lafortune et al. (2019). This value implies that capital and low-skilled labor are assumed to be substitutes. I show some suggestive evidence on the existence of capital-skill complementarity in pre-WWI Hungary in Appendix A.2.7 which corroborates this assumption. To the best of my knowledge, no estimate exists for labor's share in Austria-Hungary, so I use $\theta_{ML} = 0.5$ as my preferred value. This value lies between manufacturing's labor cost share in Sweden, which was somewhat higher in the second part of the 19th century (Bengtsson, 2012), and the Italian industrial labor share, which was lower in the decades preceding WWI

[11] See Appendix B.2 for a discussion of potential equilibria and their stability in the baseline model.

[12] Reassuringly, the yearly number of emigrants does not predict the net import of agricultural machinery in the subsequent year in the 25-30 years before WWI, controlling for a simple linear time trend (unreported).

Table 2.1: Scale elasticity estimates

θ_{ML}	σ_M	**Scale elasticity** (γ)	Note
0.5	2.44	0.101	σ_M: US 1860-1880 & 1890-1930 average, illiterates and capital; Lafortune et al., 2019 w/ Cobb-Douglas outer nest ($\rho = 0$)
0.6	2.44	0.121	θ_{ML} : Swedish manuf. c. 1900; Bengtsson, 2012
0.4	2.44	0.08	θ_{ML} : Italian industry c. 1900; Gabbuti, 2018
0.5	3.74	0.066	σ_M : US 1860-1880 & 1890-1930 average, illiterates and capital; Lafortune et al., 2019 w/ $\rho = 0.33$ outer nest
0.5	1.67	0.147	σ_M : US 1963-1992, non-college and equipment; Krusell et al., 2000
0.5	2	0.123	σ_M : physical capital and unskilled labor Stokey, 1996

Note: own calculations based on Proposition 2. The left-hand side of that expression is calculated as the ratio of the coefficients in Column 1 and 3 of Table 1.12. Then, the assumed values for θ_{ML} and σ_M are used to quantify the scale elasticity in a given row of this table. Notes in the last column explain the main difference compared to the assumptions of the first row.

(Gabbuti, 2018). Estimates with different elasticity of substitution and labor share assumptions are presented in Table 2.1.

Using these values, the elasticity estimates fall between 0.07-0.15 with my preferred estimate being 0.101. The interpretation of this scale elasticity is the following: doubling employment in an industry group-district pair is predicted to increase the total factor productivity of that industry group by 7%. This scale elasticity is similar to the average scale elasticity estimate of Bartelme et al. (2019) for two-digit manufacturing sectors at the country level in the 2000s (0.167).[13] Strong effects of intra-sector density are documented in other developing and historical settings as well. Henderson et al. (2001) demonstrate their existence at the city-manufacturing sector level during the rapid growth of South Korea. Likewise, López and Südekum (2009) find significant intra-sector spillovers in Chilean manufacturing in the 1990s. Klein and Crafts (2020a) also show that initially greater sectoral specialization in manufacturing was associated with faster subsequent productivity growth in US cities during the Second Industrial Revolution.

An obvious source of external economies of scale could be localization economies (Marshall-Arrow-Romer externalities): input-output linkages, labor pooling and knowledge spillovers. All of these channels are borne out by empirical evidence pertaining to the studied time period. First, industrial sectors tended to co-locate with their value chain partners in the US at the beginning of the 20th century (Diodato et al., 2018). Second, Kim (2006) finds that search and matching costs mattered for the emergence of industrial agglomeration forces in the US. Third, according to Baten et al. (2007), intra-industry knowledge spillovers spurred productivity growth (innovative activity) in south-western Germany around 1900. Thus, a furniture or steal maker had most likely considerable productivity gains

[13] Sector-specific values range from 0.08 to 0.42.

from a growing lumber or metal industry in its proximity.[14]

When I combine the model with industry group-level reduced form effects to quantify the scale elasticity, I implicitly assume that scale externalities exclusively stem from the own industry group. This is consistent with the evidence that the presence of inter-industry (Jacobs) externalities is not robustly estimated in the literature (Combes and Gobillon, 2015) and known to be more important for products in their development stage (Duranton and Puga, 2001). Inter-industry spillovers are likely to have been very limited in Hungary considering that Budapest is not in my sample and that sectoral diversification stimulated productivity growth merely in the largest US cities between 1880 and 1930 (Klein and Crafts, 2020a).[15]

2.7 Conclusion

The key contribution of this chapter is introducing a novel mechanism to the migration literature, explaining how low-skilled emigration can stunt local structural change in open economies. To do so, I build a two-sector small open economy model which is used to interpret the empirical facts documented for the case of pre-WWI mass emigration from Hungary. In this model, the key difference between agriculture and manufacturing lies in scale economies. In particular, when manufacturing produces under external economies of scale and it is difficult to substitute for labor in agricultural production, then emigration leads to local deindustrialization. Besides discussing the validity of the assumptions underlying the model, I also provide a point estimate on the strength of scale economies which is similar to modern estimates.

Finally, a discussion about the external validity of the results is important. I believe that a crucial difference between the underlying case of Hungary and emigration from current low-income countries lies in the availability of alternative, capital-intensive production technologies. The absence of a more extensively mechanized production method might have been more pronounced in historical time periods (e.g., see the case of mining in Abramitzky et al., 2019a) and could especially matter for agriculture. The post-WWI spread of tractors (Lew and Cater, 2018) or other labor-saving innovations (Clemens et al., 2018) might make it easier for current low-income economies to mechanize agricultural production in response

[14] An industrial inspector district consisted of four counties on average. There is evidence that scale externalities do not decay quickly with distance in other developing context (López and Südekum, 2009).

[15] A notable exception is Hanlon and Miscio (2017) who present evidence on significant inter-industry effects. While Klein and Crafts (2020a) analyze labor productivity growth in manufacturing in US cities, Hanlon and Miscio (2017) examine employment growth in a broader set of sectors in the largest cities of the United Kingdom.

to emigration. Nevertheless, modern technology adoption might not go seamlessly. It has been established that the adoption of predecessor technologies can foster modern technology adoption (Comin and Hobijn, 2004); however, Comin et al. (2008) show that technology usage lags more than one hundred years behind the US in many developing countries and more than 180 years in most Sub-Saharan African countries. Thus, lack of predecessor technology adoption may be a barrier to modern technology adoption, in addition to (human) capital shortages. Another major difference between pre-WWI Hungary and current developing countries concerns infrastructure. Hungary had high railway density even before the First World War (70km per 1.000 km^2; Kaposi, 2002), making local markets relatively open. This figure is only 19.4 in Bangladesh, 1.5 in Botswana, 2.1 in Cameroon, 0.7 in Ethiopia or 20.3 in India, according to recent World Bank data (2016), suggesting that local structural change may rely more on local demand in these economies than in Austria-Hungary (Emerick, 2018; Santangelo, 2019). Furthermore, while the role of remittances is not explicitly analyzed in this work owing to data limitations, they are implicitly included in the estimation. Puskás (1982) estimated remittances to be of substantial size at the aggregate level, comparable to modern low-income countries. Thus, a potential absence of remittance inflows is unlikely to make a difference between the studied time period and current cases of emigration. In conclusion, further research pinning down the conditions under which my proposed economic mechanism might dominate within-sector capital-intensive technology adoption is much needed.

B Appendix - Chapter 2

B.1 A literature survey of agglomeration economies in historical contexts

There are numerous pieces of evidence supporting the existence of (industrial) agglomeration externalities in many countries in the late nineteenth and early twentieth centuries. Studying the regional distribution of economic activity in the US, Kim (1995) presents supporting evidence on scale economies and the importance of relative endowments between 1860 and 1987. Focusing on Spain in the 19$^{\text{th}}$ century, Rosés (2003) finds similar forces as Kim (1995). Also for Spain, Martínez-Galarraga et al. (2008) estimate a positive effect of the density of economic activity on labor productivity in industry, starting in the mid-nineteenth century. In France, Combes et al. (2011) find considerable agglomeration economies since the 1860s (their first observations). Agglomeration economies played a major role even in the rural periphery of post-WWII Finland (Sarvimäki, 2011). Martínez-Galarraga (2012) demonstrates the role of scale economies and factor endowments (especially local labor supply) in determining the spatial distribution of economic activity as early industrialization progressed in Spain. Analyzing the case of the Tennessee Valley Authority in the 1930s, Kline and Moretti (2014) find evidence on agglomeration economies in manufacturing. For the case of Russia, Buggle and Nafziger (2021) suggest that early industrial development and subsequent agglomeration effects can be important channels of persistence of the effects of historical serfdom. Additionally, Lafortune et al. (2021) argue that previous estimates on increasing *internal* returns to scale in manufacturing in the US during the Second Industrial Revolution are substantially upward biased and confounded by strong agglomeration effects, especially before 1900. Finally, Peters (2021) shows that places, where ethnic Germans were transferred to from Eastern Europe following WWII, experienced a persistent, positive labor supply shock, resulting in higher income and manufacturing employment per capita over time.

Interestingly, even Hungarians realized the importance of scale externalities in the period of mass migration: «*why are our [i.e. Hungarian] factories not competitive? We have manufacturing, but we lack the* ***atmosphere*** *of manufacturing... and we will only have this* ***atmosphere*** *once many factories will exist. What few factories cannot do, many factories can*» (MGYOSZ, 1907, p. 342-343). This sentence also bears some resemblance to the «big push» argument (Murphy et al., 1989), namely entering factories can provide non-internalized incentives for other firms to enter. This may be relevant during the Second Industrial Revolution as well (Mokyr and Voth, 2010).

B.2 Equilibrium stability and uniqueness

In this section, I firstly write down regularity conditions which guarantee the existence of at least one stable internal equilibrium in the sense of Krugman (1991). Then, I provide an additional discussion about the uniqueness of this equilibrium.

First, to guarantee the existence of at least one (stable or unstable) internal equilibrium, I can only write a very general condition because the two CES production functions do not allow me to derive a closed-form expression.

Proposition 6. *If there exists an $x \in (0, \overline{L})$ such that $w_A(x) - w_M(\overline{L} - x) \leq 0$, then there is at least one internal equilibrium.*

Relying on graphical intuition, this proposition guarantees that w_A and w_M cross at least once, implying an internal equilibrium in which the agricultural and manufacturing wage levels are equal (note that $w_A(\overline{L}) > w_M(0)$). This proposition rules out the case when only the fully agricultural, preindustrial equilibrium exists. Figure B1 shows that, in the latter case, the manufacturing wage curve is below the agricultural one for all levels of agricultural employment, providing no incentives for agricultural workers to move to manufacturing. This agriculture-only case

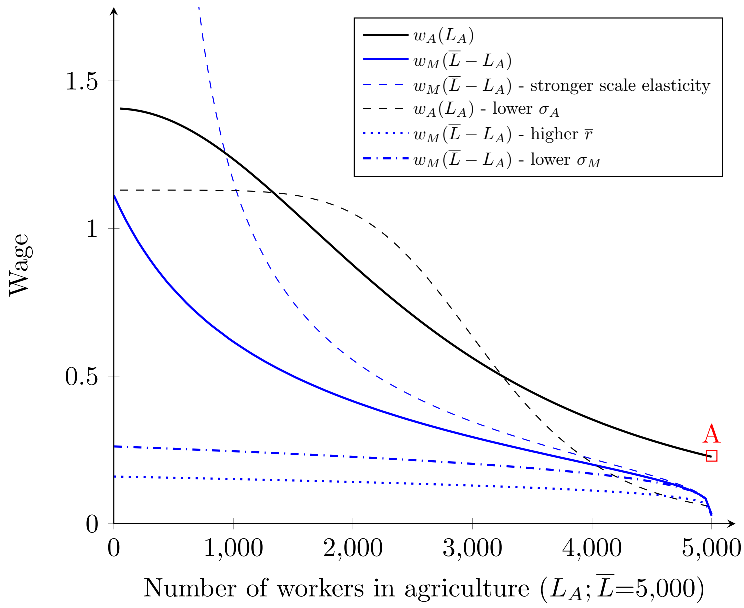


Figure B1: The pre-industrial equilibrium is the only equilibrium

is less likely to hold if industrial scale economies are stronger or the elasticity of substitution is lower in agriculture. On the other hand, a higher interest rate or lower elasticity of substitution in manufacturing make a single, preindustrial equilibrium more likely.

Agriculture always offers a positive wage, even if $L_A = \overline{L}$, when manufacturing has a wage of zero (and a capital stock of zero). Consequently, the agriculture-only equilibrium is always stable. However, this implies that, if there is a single internal equilibrium, it must be unstable because w_M crosses w_A from below. Thus, the marginal agricultural worker has an incentive to move to manufacturing as this worker would earn an even higher wage there, making this equilibrium unstable. To guarantee that there is at least one stable internal equilibrium, the following sufficient condition can be written which essentially limits the strength of scale economies.

Proposition 7. *If γ is such that $(1-\beta)^{\frac{\sigma_A}{\sigma_A-1}} = w_A(0) > w_M(\overline{L}) = \left[\frac{(\overline{L}^{\gamma}\overline{p}_M)^{1-\sigma_M} - \alpha^{\sigma_M}\overline{r}^{1-\sigma_M}}{(1-\alpha)^{\sigma_M}}\right]^{\frac{1}{1-\sigma_M}}$, then there is a stable internal equilibrium.*

In practice, an upper bound can be calculated for the scale elasticity as a function of other exogenous parameters ($\overline{L}, \sigma_A, \sigma_M$, etc.). Using the graphical intuition again, this condition guarantees that the w_M curve crosses the vertical axis below the w_A curve. Therefore, if Proposition 6 also holds, the two wage curves cross at least twice, the second equilibrium being a stable one. Consequently, the

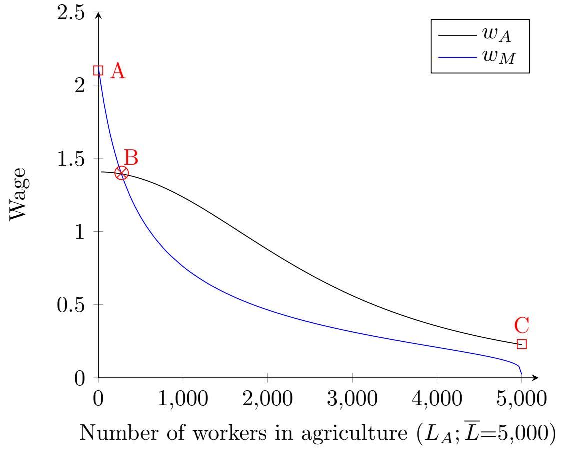


Figure B2: Corner equilibria are the stable equilibria

case shown in Figure B2 (only 'A' and 'C', the two corner solutions, are stable equilibria) can be ruled out if this proposition holds.

Additionally, using the second derivatives of $w_A(L_A)$ and $w_M(L_M)$, it is easy to demonstrate that both curves have at most one inflection point on $[0; \overline{L}]$. Thus, there may be at most three crossings of the two wage curves (the third being an unstable equilibrium). The case consisting every possible equilibrium is shown in Figure B3. Equilibria 'A' and 'E' are stable corner solutions, while 'C' is the stable internal equilibrium. In conclusion, if a stable internal equilibrium exists, it must be unique. In addition, notice that the stable internal equilibrium satisfies the strong complementarity condition ($\partial \hat{w} / \partial \hat{\overline{L}} > 0$),[16] while the unstable internal equilibria never do so.

B.3 Equilibrium conditions following Jones (1965)

To analyze how an exogenous change in the local labor endowment affects the initial equilibrium, I first differentiate the profit maximizing conditions which results in:

$$c_M(w, \overline{r}, v) = \overline{p}_M \Longrightarrow \theta_{ML} \cdot \hat{w} = \hat{v} \tag{2.8}$$

$$c_A(w, r_T) = 1 \Longrightarrow \theta_{AL} \cdot \hat{w} + \theta_{AT} \cdot \hat{r}_T = 0 \tag{2.9}$$

[16] In other words, the equilibrium wage and the number of industrial workers drop following emigration (a leftward shift of w_M).

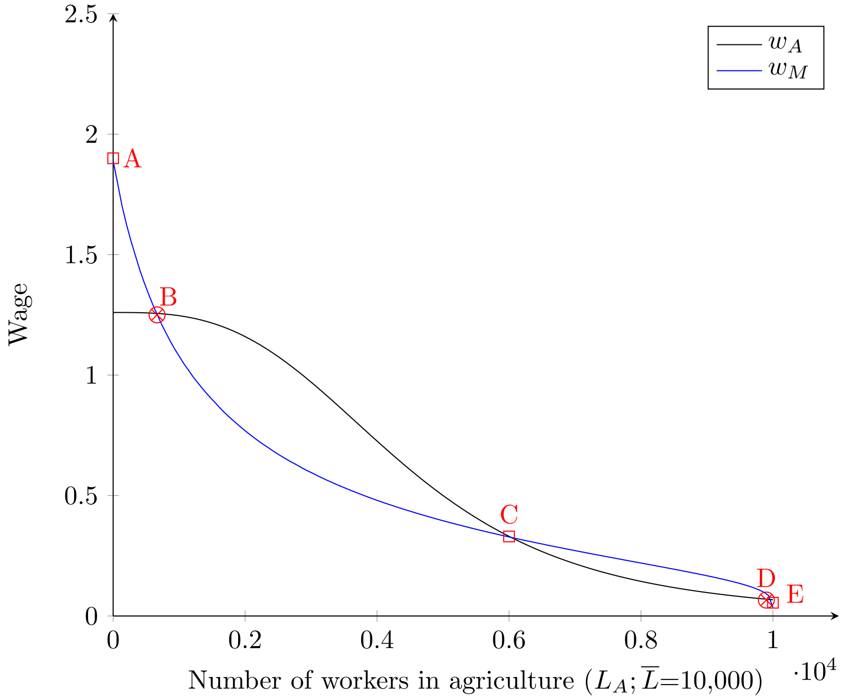


Figure B3: Equilibrium stability and uniqueness

where θ_{ij} represents the cost share of factor j in sector i. Differentiating the market clearing conditions leads to two additional equations:

$$\overline{L} = a_{ML} \cdot Y_M + a_{AL} \cdot Y_A \Longrightarrow \hat{\overline{L}} = \lambda_{ML} \cdot (\hat{a}_{ML} + \hat{Y}_M) + \lambda_{AL} \cdot (\hat{a}_{AL} + \hat{Y}_A) \tag{2.10}$$

$$\overline{T} = a_{AT} \cdot Y_A \Longrightarrow \hat{a}_T + \hat{Y}_A = 0 \tag{2.11}$$

where λ_{iL} denotes the initial share of labor employed in sector i and the exogenous change of interest is $\hat{\overline{L}}$. The detailed derivation of the previous four equations is shown in the next section below.

Following Jones (1965), I have two additional equilibrium conditions as a constant elasticity of substitution connects changes in unit factor demands with changes in input prices:

$$\sigma_M = \frac{\hat{a}_{MK} - \hat{a}_{ML}}{\hat{w} - \hat{\overline{r}}} = \frac{\hat{a}_{MK} - \hat{a}_{ML}}{\hat{w}} \tag{2.12}$$

$$\sigma_A = \frac{\hat{a}_{AT} - \hat{a}_{AL}}{\hat{w} - \hat{r_T}}. \tag{2.13}$$

If firms treat input prices as fixed, the following conditions must also hold:[17]

$$\theta_{ML} \cdot \hat{a}_{ML} + \theta_{MK} \cdot \hat{a}_{MK} = 0 \tag{2.14}$$

$$\theta_{AL} \cdot \hat{a}_{AL} + \theta_{AT} \cdot \hat{a}_{AT} = 0. \tag{2.15}$$

The last equation is simply the log-linearization of v:

$$\hat{v} = \gamma \cdot \hat{L}_M = \gamma \cdot \hat{a}_{ML} + \gamma \cdot \hat{Y}_M. \tag{2.16}$$

With nine endogenous variables ($\hat{a}_{ML}, \hat{a}_{MK}, \hat{a}_{AL}, \hat{a}_{AT}, \hat{Y}_M, \hat{Y}_A, \hat{v}, \hat{w}, \hat{r}_T$) and nine independent equations (Eq. 2.8-2.16), every endogenous variable can be expressed as a function of exogenous parameters, including $\hat{\overline{L}}$.

B.4 Detailed derivation of the equilibrium conditions

The differentiation of the two profit maximization conditions yields:

$$\begin{aligned}
dc_M(w, \overline{r}, v) &= d\overline{p}_M = 0 \\
\frac{\partial c_M(w, \overline{r}, v)}{\partial w} \cdot dw + \frac{\partial c_M(w, \overline{r}, v)}{\partial \overline{r}} \cdot d\overline{r} + \frac{\partial c_M}{\partial v} \cdot dv &= 0 \\
a_{ML} \cdot dw + c_M(w, \overline{r}) \cdot \frac{-1}{v^2} \cdot dv &= 0 \\
a_{ML} \cdot dw - \frac{c_M(w, \overline{r})}{v} \cdot \frac{1}{v} \cdot dv &= 0 \\
a_{ML} \cdot dw - \overline{p}_M \cdot \frac{dv}{v} &= 0 \\
\frac{a_{ML} \cdot w}{\overline{p}_M} \cdot \frac{dw}{w} - \frac{dv}{v} &= 0 \\
\theta_{ML} \cdot \hat{w} &= \hat{v}
\end{aligned}$$

$$\begin{aligned}
dc_A(w, r_T) &= d\overline{p}_A = 0 \\
\frac{\partial c_A}{\partial w} \cdot dw + \frac{\partial c_A}{\partial r_T} \cdot dr_T &= 0 \\
a_{AL} \cdot dw + a_{AT} \cdot dr_T &= 0 \\
\frac{a_{AL} \cdot w}{\overline{p}_A} \cdot \frac{dw}{w} + \frac{a_{AT} \cdot r_T}{\overline{p}_A} \cdot \frac{dr_T}{r_T} &= 0 \\
\theta_{AL} \cdot \hat{w} + \theta_{AT} \cdot \hat{r}_T &= 0
\end{aligned}$$

[17] For example, Eq. 2.14 can be derived as follows. Differentiate $p_m = w \cdot a_{ML} + \overline{r} \cdot a_{MK}$ to get $0 = w \cdot da_{ML} + \overline{r} \cdot da_{MK}$. Finally, divide both sides by p_M, and divide and multiply da_{Mj} by a_{Mj}.

The differentiation of the labor market clearing condition yields:
$d\overline{L} = da_{ML} \cdot Y_M + a_{ML} \cdot dY_M + da_{AL} \cdot Y_A + a_{AL} \cdot dY_A = \frac{da_{ML}}{a_{ML}} \cdot a_{ML} \cdot Y_M + a_{ML} \cdot Y_M \cdot \frac{dY_M}{Y_M} + \frac{da_{AL}}{a_{AL}} \cdot a_{AL} \cdot Y_A + a_{AL} \cdot Y_A \cdot \frac{dY_A}{Y_A} = \hat{a}_{ML} \cdot a_{ML} \cdot Y_M + \hat{Y}_M \cdot a_{ML} \cdot Y_M + \hat{a}_{AL} \cdot a_{AL} \cdot Y_A + \hat{Y}_A \cdot a_{AL} \cdot Y_A = (\hat{a}_{ML} + \hat{Y}_M) \cdot a_{ML} \cdot Y_M + (\hat{a}_{AL} + \hat{Y}_A) \cdot a_{AL} \cdot Y_A$

Divide both sides by $\overline{L}$ to arrive at:
$\hat{\overline{L}} = \lambda_{ML} \cdot (\hat{a}_{ML} + \hat{Y}_M) + \lambda_{AL} \cdot (\hat{a}_{AL} + \hat{Y}_A)$

B.5 Proof of Propositions 1 and 2

First, combine the log-linearized scale externality with Eq. 2.8:
$\hat{v} = \gamma \cdot \hat{a}_{ML} + \gamma \cdot \hat{Y}_M = \theta_{ML} \cdot \hat{w}$.

Dividing both sides by γ gives an expression for $\hat{a}_{ML} + \hat{Y}_M$, which is used in the log-linearized labor market clearing condition:
$\hat{\overline{L}} = \lambda_{ML} \cdot (\hat{a}_{ML} + \hat{Y}_M) + \lambda_{AL} \cdot (\hat{a}_{AL} + \hat{Y}_A) = \frac{\lambda_{ML} \cdot \theta_{ML}}{\gamma} \cdot \hat{w} + \lambda_{AL} \cdot (\hat{a}_{AL} - \hat{a}_{AT})$

where the latter follows from the log-linearized land market clearing condition: $-\hat{a}_{AT} = \hat{Y}_A$. $\hat{a}_{AL} - \hat{a}_{AT}$ can be further re-written using the elasticity of substitution in agriculture (Eq. 2.13). Thus, we have:
$\hat{\overline{L}} = \left(\frac{\lambda_{ML} \cdot \theta_{ML}}{\gamma} - \lambda_{AL} \cdot \sigma_A\right) \cdot \hat{w} + \lambda_{AL} \cdot \sigma_A \cdot \hat{r}_T$.

Combining this expression with $\theta_{AL} \cdot \hat{w} + \theta_{AT} \cdot \hat{r}_T = 0$ yields:
$\hat{\overline{L}} = \frac{\lambda_{ML} \cdot \theta_{ML} \cdot \theta_{AT} - \lambda_{AL} \cdot \sigma_A \cdot \theta_{AT} \cdot \gamma - \lambda_{AL} \cdot \sigma_A \cdot \theta_{AL} \cdot \gamma}{\gamma \cdot \theta_{AT}} \cdot \hat{w}$.

Noticing that $\theta_{AT} + \theta_{AL} = 1$ simplifies the expression to:
$\hat{\overline{L}} = \frac{\lambda_{ML} \cdot \theta_{ML} \cdot \theta_{AT} - \lambda_{AL} \cdot \sigma_A \cdot \gamma}{\gamma \cdot \theta_{AT}} \cdot \hat{w}$.

Rearranging terms leads to the first equality of Proposition 1. In order to get the second equality, I derive employment changes in the two sectors.
$\hat{L}_M = \hat{a}_{ML} + \hat{Y}_M = \frac{\theta_{ML}}{\gamma} \cdot \hat{w}$

where the second equality follows from the relationship between changes in the scale externality and wages, already used in this proof.
$\hat{L}_A = \hat{a}_{AL} + \hat{Y}_A = \hat{a}_{AL} - \hat{a}_{AT} = \sigma_A \cdot (\hat{r}_T - \hat{w}) = -\sigma_A \cdot \left(\frac{\theta_{AL}}{\theta_{AT}} + 1\right) \cdot \hat{w} = -\frac{\sigma_A}{\theta_{AT}} \cdot \hat{w}$

where the first equality follows from the log-linearized land market clearing condition, the second equality from Eq. 2.13, the third equality from Eq. 2.15. One must also notice that $\theta_{AT} + \theta_{AL} = 1$.

Finally, the change in the capital stock can be expressed as a function of $\hat{w}$:
$\hat{K} = \hat{a}_{MK} + \hat{Y}_M = \hat{a}_{MK} + \frac{\theta_{ML}}{\gamma} \cdot \hat{w} - \hat{a}_{ML} = \left(\sigma_M + \frac{\theta_{ML}}{\gamma}\right) \cdot \hat{w}$

where the second equality follows from combining the earlier derived expression for $\hat{L}_M$ with $\hat{L}_M - \hat{a}_{ML} = \hat{Y}_M$, and the third one from Eq. 2.12. All the expressions are derived to prove Proposition 2, too.

B.6 Proof of Proposition 3

The log-differentiated form of the modified labor market clearing condition is the following:
$\hat{\bar{L}} = \lambda_{ML_1} \cdot (\hat{a}_{ML_1} + \hat{Y}_{M_1}) + \lambda_{ML_2} \cdot (\hat{a}_{ML_2} + \hat{Y}_{M_2}) + \lambda_{AL} \cdot (\hat{a}_{AL} + \hat{Y}_A) = \frac{\theta_{ML_1}}{\gamma_1} \cdot \hat{w} + \frac{\theta_{ML_2}}{\gamma_2} \cdot \hat{w} + \lambda_{AL} \cdot (\hat{a}_{AL} + \hat{Y}_A)$ where the second equality follows from the relationship between the log-linearized scale externality and wages in each sector of manufacturing: $\hat{v}_i = \gamma_i \cdot \hat{a}_{ML_i} + \gamma_i \cdot \hat{Y}_{M_i} = \theta_{ML_i} \cdot \hat{w}$. After this point, the same steps are needed as in the proof of Proposition 1.

B.7 Proof of Proposition 5

Similarly to the earlier derivation, the elasticity of substitution connects factor price changes with changes in the unit factor demand:

$$\sigma_1 = \frac{\hat{a}_{AT} - \hat{a}_X}{\hat{p}_X - \hat{r_T}},$$
$$\sigma_2 = \frac{\hat{a}_{AK} - \hat{a}_{AL}}{\hat{w}}.$$

It is easy to obtain linear approximations for $\hat{p}_X$ and $\hat{a}_X$ using the following expressions implied by the CES form of the production function:

$$p_X = \left[w^{1-\sigma_2} + \bar{r}_{K_A}^{1-\sigma_2}\right]^{\frac{1}{1-\sigma_2}} \implies \hat{p}_X = \left(\frac{w}{p_X}\right)^{1-\sigma_2} \cdot \hat{w} = \delta_w \cdot \hat{w}$$

$$a_X = \left[a_{AL}^{\frac{\sigma_2-1}{\sigma_2}} + a_{AK}^{\frac{\sigma_2-1}{\sigma_2}}\right]^{\frac{\sigma_2}{\sigma_2-1}} \implies \hat{a}_X = \left(\frac{a_{AL}}{a_X}\right)^{\frac{\sigma_2-1}{\sigma_2}} \cdot \hat{a}_{AL} + \left(\frac{a_{AK}}{a_X}\right)^{\frac{\sigma_2-1}{\sigma_2}} \cdot \hat{a}_{AK} = \delta_L \cdot \hat{a}_{AL} + \delta_K \cdot \hat{a}_{AK}.$$

Combining the previous four equations with $\theta_{AK} \cdot \hat{a}_{AK} + \theta_{AL} \cdot \hat{a}_{AL} + \theta_{AT} \cdot \hat{a}_{AT} = 0$ and $\frac{\theta_{AL}}{\theta_{AT}} \cdot \hat{w} = -\hat{r}_T$ yields the expression in the proposition.

Chapter 3

THE MECHANICS OF GOOD FORTUNE

ON INTERGENERATIONAL MOBILITY DURING THE SECOND INDUSTRIAL REVOLUTION

3.1 Introduction

Structural transformations have a profound impact on the career and socio-economic status of most people. In particular, recent waves of robotization or trade shocks changed the structure of the labor market and the life of millions of workers depending on their industry or occupation (Acemoglu and Restrepo, 2019, 2020; Autor et al., 2016; Dauth et al., 2021a,b; Graetz and Michaels, 2018; Humlum, 2019; Traiberman, 2019). However, since these shocks are very recent and the grandchildren of affected workers have not even been born yet, we can merely speculate how the offspring of demanded tech workers or of displaced manufacturing workers might fare in the very long run. Therefore, in this paper we go back in time to study the effect of an arguably equally disrupting time period on the labor market: the Second Industrial Revolution (ca. 1870-1914). We try to understand *to what extent* and *how* members of a particularly demanded occupation - machinists[1] - could pass on their gains in socio-economic status to later generations in the United States.

This chapter is co-authored with Laurenz Bärtsch.

[1]Workers in charge of installing and maintaining machinery.

To the best of our knowledge, this is the first work which documents the persistence in income gains caused by a labor market shock on the grandchildren of affected individuals, i.e. over three generations. Moreover, we also shed light on the mechanisms which underlie the documented intergenerational persistence: internal migration and increased (secondary) education. We show that these two channels may account for the entire positive effect of machinists on their offspring's earnings, though their relative importance depends on initial urban status: the offspring of initially rural machinists gained from more schooling as well as from internal migration to more urban areas, whereas the channel of internal migration does not play a significant role for the sons of urban machinists.

Using the US full count census, we can overcome the main hurdle to intergenerational studies: the scarcity of data connecting generations. We exploit this data set leveraging the strengths of two, complementary estimation methods: propensity score matching and fixed effects regression. Our empirical strategy amounts to comparing the post-1870 outcomes of machinists to those of non-machinists, who were observationally very similar to machinists before the onset of the Second Industrial Revolution. Next, we identify the offspring of these individuals and investigate their outcomes. In our baseline strategy, we use personal and residential characteristics from the census as controls and complement them with occupation-based education (Song et al., 2020) and novel earnings scores. These earnings scores, constructed based on U.S. Department of Labor (1900), are another contribution of this paper as we are the first ones to calculate state-specific earnings scores for a large number of occupations before 1890.

In this paper, we document that machinists could pass on their relative gains in socio-economic status to their (grand)sons. First, we find that machinists, whose occupation experienced a relative labor demand boom starting in the 1870s, enjoyed higher earnings and occupational stability, and were more likely to live in urban places after 1870. As explained in Section 3.2, the surge in demand for machinists resulted from innovations leading to mechanization and the rapid spread of factory production methods in the US. Therefore, much demanded machinists could avoid switching to lower-paying, often agricultural occupations during the volatile business cycle of the Gilded Age. Thus, besides a relative wage improvement, the identified occupational earnings gains are driven by less occupational downgrading rather than occupational upward mobility, which could be suggestive of unobserved ability. Second, the sons of machinists held occupations with 5-12 log-points higher real or nominal earnings scores than the sons of comparable non-machinists in 1900.[2] Finally, a positive effect is estimated on the individual- or occupation-level income of grandsons in 1940, seventy years after 1870.

Next, we shed light on the mechanisms behind the documented intergenera-

[2]After correcting for the bias which stems for mismeasurement. See Section 3.6.4.

tional transmission. For the sons of initially rural machinists, the positive earnings effect partly stems from a higher probability of living in an urban area as an adult (urban wage premium). To quantify the approximate size of earnings gains originating from rural-to-urban migration, we multiply the differential likelihood of urban status with an earnings score-based estimate of the rural-to-urban migration premium pertaining to the early-twentieth-century United States (Ward, forthcoming).

Additionally, machinists' sons benefited from parental investment in their education irrespective of initial urban status, receiving approx. 0.35 more years of (mainly secondary) schooling. To study the role of education in explaining the earnings effect, we simply combine our years of schooling point estimate with a returns to schooling estimate in Goldin and Katz (2000). Moreover, by exploiting a newly digitized, county-level data set on high school provision, we establish that the positive earnings and especially schooling effects on the sons of machinists increased in county-level private high school provision.[3] This complementarity between a machinist father and local high school supply was especially strong when free-of-charge public high school supply was limited and private schools had a high teacher-student ratio. On the other hand, gains from private high schools decreased if these schools could be attended at a low price and put emphasis on scientific education in their curricula (e.g., mechanical drawing), the type of knowledge which the sons of machinists could more easily acquire at home. This suggests that passing on scientific knowledge in an informal way, within a family also helped machinist's sons succeed. Furthermore, the estimated positive effects on machinists' sons declined in public high school provision as well. This empirical result is consistent with financially more constrained non-machinist parents (Becker and Tomes, 1979, 1986). Last, we estimate a coefficient on education which is not significantly different from zero for sons who were older than ten years in 1870, suggesting that machinists were not differentially more likely to invest in the education of their sons before 1870.[4]

Apart from heterogeneity exercises, we conduct a series of robustness checks to mitigate concerns that the identified positive effects can be explained by (the transmission of) the machinists' unobserved ability. Arguably the most convincing robustness checks are regressions containing family-fixed effects, i.e. comparing machinists to their own brothers.[5] The results from this specification, which con-

[3]We show that this effect is driven by the medium-level tuition fee. At this cost level, education was less affordable for rival boys but not prohibitively costly for machinists.

[4]In accordance with the literature documenting dynamic complementarities in the production of human capital (see Heckman and Cunha, 2007), it was arguably already too late to invest in their education when the relative earnings of machinists started to rise.

[5]This robustness test can only be conducted for the generation of machinists themselves because we run into sample size limitations for later generations.

trols for the similar environment of upbringing and inherited genes, are qualitatively and quantitatively similar to those obtained from the baseline analysis. In addition, the lack of correlation - both within and across families - between the machinist indicator and standard (historical) proxies of unobserved ability (e.g., number of children or spousal literacy - measured in 1870) suggests that machinist fathers were not more able compared to their brothers or comparable peers.

We demonstrate that the results are not driven by occupation-state or census division-level pre-trends (e.g., changes in the employment share, probability of switching to agriculture, etc. in the 1850s and 1860s), and are insensitive to which specific occupations are the dominant «control occupations». Moreover, our preferred propensity score matching strategy eliminates initially large differences in the overwhelming majority of characteristics of wives, fathers and next-door-neighbors between machinists and non-machinists - even *without* matching on these characteristics. We also establish that similarly aged sons of younger and older machinists experienced similar positive effects, indicating the absence of early sorting into the machinist occupation by more talented individuals. Additionally, the inclusion of birth state-destination county (1870)-fixed effects makes it very unlikely that the results reflect spatial sorting prior to 1870.

Related literature This work is closely connected to the literature which examines the effect of parental labor market shocks on affected children. Exploiting layoffs, Hilger (2016) and Mörk et al. (2020) find at most very small negative effects on the education and adult earnings of affected children.[6] As both papers point out, these might be the consequence of a generous welfare state offsetting otherwise reduced parental spending on education. A more accurate comparison to our setting might come from papers that focus on less developed countries with a rather weak welfare state or low-income (financially constrained) families. These papers tend to find that changes in parental income - not necessarily induced by job loss - do matter for the offspring (see, e.g., Aizer et al., 2016; Akee et al., 2010; Dahl and Lochner, 2012; Di Maio and Nisticò, 2019; Løken et al., 2012; Manoli and Turner, 2018). Surveying the literature, Cooper and Stewart (2017) conclude that there is «*strong evidence that income has causal effects on a wide range of children's outcomes, especially in households on low incomes*», whereas wealth shocks do not seem to have substantial effects on children either in a historical (Bleakley and Ferrie, 2016) or in a modern context (Cesarini et al., 2017). Addi-

[6]Early papers tend to exploit mass layoffs or factory closures, and find mixed effects on schooling and future earnings of children affected by parental job loss (Bratberg et al., 2008; Coelli, 2011; Oreopoulos et al., 2008; Rege et al., 2011). However, Hilger (2016) argues that many early findings on large, negative effects might be driven by the assortative matching of low-quality workers and low-quality firms leading to selection into layoffs or closure. Løken (2010), exploiting the oil boom in Norway as a permanent income shock, finds no effect on children either.

tionally, there is a large literature documenting the role of credit constraints and grants, mostly for college education.[7] Our contribution is to show that the effect of labor market shocks may persist even for the offspring in the second generation. In addition, we pin down mechanisms which lead to the documented intergenerational persistence. These are not well-understood even in the modern context and, to the best of our knowledge, have not been studied in a historical context yet.

This paper also speaks to a literature which seeks to identify the determinants of intergenerational mobility in the 19^{th}-20^{th} century United States. Parman (2011) demonstrates that children from high-income families benefited disproportionally more from improving public high school availability in Iowa at the turn of the 20^{th} century, resulting in a higher intergenerational income elasticity. However, we find a negative association between public high school supply and the relative gains of machinists' sons, in line with Solon (2004) and Olivetti and Paserman (2015). Since parents of similar socio-economic background tend to have similar preferences over education (Boneva and Rauh, 2018), we believe that comparing machinist fathers to fathers in other middle-class occupations might eliminate the effect uncovered by Parman (2011) in our case. In a comparison of migrating to non-migrating brothers, Ward (forthcoming) finds that rural-urban migration was an important contributor to upward mobility in the early-twentieth-century US, particularly so for people from the poorest households. This finding is in line with our results on the importance of urban place of living for initially rural machinists' sons. Furthermore, Olivetti and Paserman (2015) and Song et al. (2020) show that industrialization was a major determinant of a relatively low intergenerational mobility around 1900. Our case study of machinists aligns well with this view and suggests highly persistent positive effects on their offspring.

By analyzing the effect of a change in occupational labor demand on machinists themselves, this work is also connected to a fast growing literature which investigates the effect of technology-induced occupational labor demand changes on affected individuals. Papers studying the impact of automation or robotization typically find that robots decrease the employment share of lower-skilled production workers and benefit workers in occupations with complementary tasks - just as early machines did to machinists (Acemoglu and Restrepo, 2020; Dauth et al., 2021b; Graetz and Michaels, 2018; Humlum, 2019). Focusing on the automation of telephone operation, Feigenbaum and Gross (2020) find that incumbent telephone operators bore most of the losses: they were more likely to be in lower-paying occupations or left the labor force entirely after automation started. However,

[7] There is ample evidence that credit constraints and grants for schooling matter even in modern contexts and in many developed countries. The early literature is summarized in Lochner and Monge-Naranjo (2012), see also Bettinger et al. (2019), Castleman and Long (2016), Denning et al. (2019), Fack and Grenet (2015), Hai and Heckman (2017), Lee and Seshadri (2019), Molina and Rivadeneyra (2021), Solis (2017), and Wright (2021) .

growth in middle-skill jobs absorbed the labor supply of later generations. Using exceptionally disaggregated Swedish data on occupations, Edin et al. (2019) show that those facing occupational decline lost about 2-5 percent of mean cumulative earnings and were less likely to remain in their starting occupations - the mirror image of what we estimate in the US for machinists. Additionally, Swedish earnings losses are partly accounted for by reduced employment and increased time spent in unemployment and retraining. Our contribution to this literature lies in analyzing a different time period, mainly the Second Industrial Revolution, in detail.

The paper is structured as follows. First, Section 3.2 discusses the historical background, then, Section 3.3 addresses questions related to data sources and sample construction. Section 3.4 presents the empirical strategy while Section 3.5 contains the main results. Thereafter, the reader may find a battery of robustness exercises and a discussion of a non-classical measurement error in Section 3.6. Finally, Section 3.7 concludes.

3.2 Historical background

The machinist occupation was born in the First Industrial Revolution in the United Kingdom, but members of this occupation played an important role in innovative activities in the United States in the early nineteenth century as well (Kelly et al., 2020; Meisenzahl and Mokyr, 2011; Sokoloff and Khan, 1990). Nevertheless, professional engineers had taken over this inventive role by the mid-nineteenth century, even before the Second Industrial Revolution started (Hanlon, 2021; Maloney and Valencia Caicedo, 2020). Thus, the assembly and maintenance of industrial machinery was left as the task of most machinists (U.S. Department of Labor, 1899). People could enter this occupation through the helper system, a type of informal apprenticeship. This meant initially simple operations followed by a sequence of more demanding tasks as they gained experience next to senior machinists. Additionally, the division of labor among American machinists reached a substantially higher level compared to the UK, resulting in a relatively lower skill requirement and making a cross-country earnings comparison of machinists almost impossible (Rosenbloom, 2002). In spite of reduced skill requirements in the US, machinists remained a part of the so-called «labor aristocracy» alongside other skilled craftsmen, for instance, blacksmiths, carpenters, conductors, masons, painters or plumbers (Dawson, 1979; Rosenbloom, 2002).

While at-scale factory production was limited to the textile industry until the Civil War, the situation changed rapidly after the onset of the Second Industrial Revolution around 1870. Mechanization and factory production methods spread swiftly across a wide range of industries, led by steel and chemicals production, and was supercharged by the utilization of electricity and novel ways of transportation

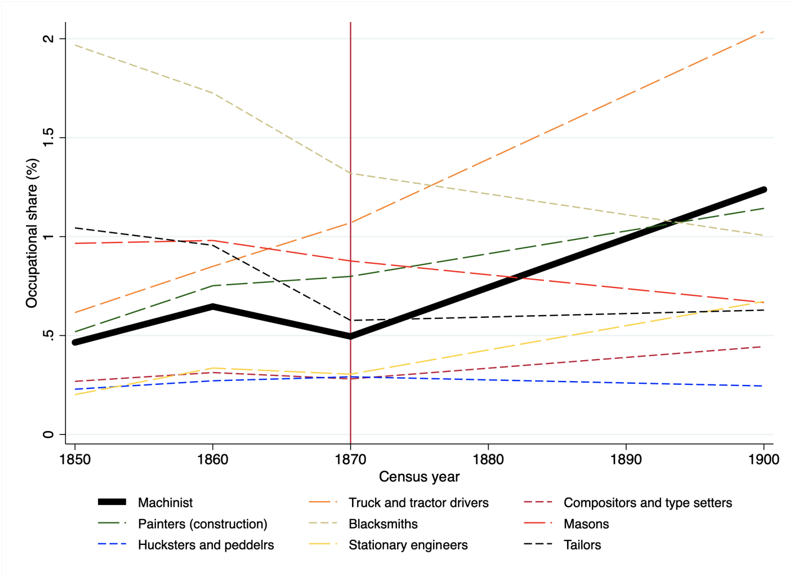


Figure 3.1: The evolution of occupational employment shares over time

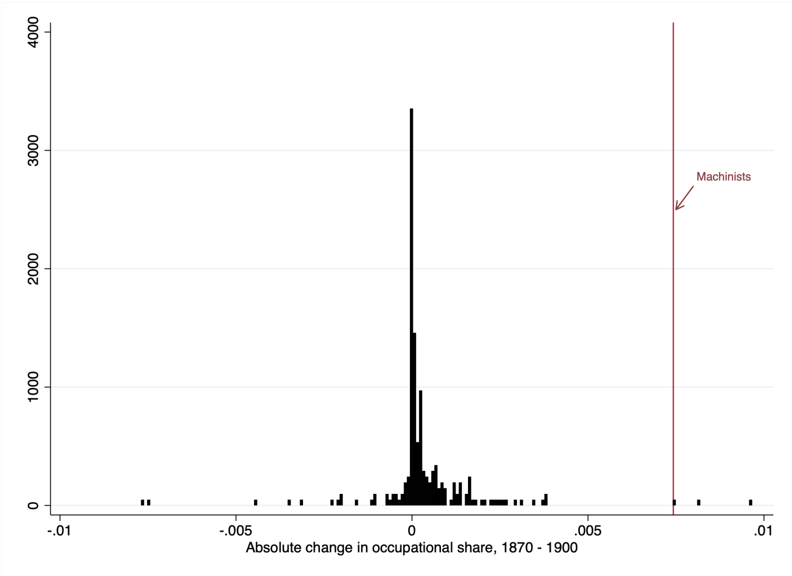


Figure 3.2: Histogram of occupational employment changes

Notes: **Figure 1**: the sample includes all males aged 16-65 who did not give a non-occupational response in the full count census in a given year. The number of workers in each occupation is divided by the total number of workers in 1850, 1860, 1870 and 1900. Harmonized occupations (1950) are used. Therefore, people classified as 'Truck and tractor drivers' were predominantly teamsters in the nineteenth century. **Figure 2**: the same sample used as in Figure 1. Only the employment of 'Mine operatives and laborers' and 'Truck and tractor drivers' grew faster than that of machinists out of the narrowly defined (i.e., not 'not elsewhere classified') occupations (see Section 3.3.1).

(e.g., the railway; Mokyr, 1999b; Rosenbloom, 2002). American manufacturing harnessed steam engines with a total capacity of approx. 1.000 thousand HP in 1870. This figure exponentially increased to almost 9.000 thousand by 1900 (Rosenberg and Trajtenberg, 2004). The average establishment size stagnated between 1849 and 1869, but it experienced a historically unprecedented growth in the 1870s and 1880s as production concentrated in factories (O'Brien, 1988).

As assembling, setting up and maintaining the machinery were the main tasks of machinists, they were found in a wide range of mechanized sectors by the end of the 19th century (U.S. Department of Labor, 1899): brush, buttonmold, canned corn, cigarette, faucet, female shoes, ingrain carpet, needle or teaspoon production, etc.. The sudden need for expertise to handle machines in these sectors led to a fast rising demand for machinists. The change in their employment share, which could barely outpace the growth of the labor force and was similar to that of some other craftsmen prior to 1870, experienced a steep acceleration (see Figure 3.1). As a result, their number almost doubled between 1870 and 1880, and a five-fold increase is registered in the full count census between 1870 and 1900. The US population merely doubled in these three decades. Thus, the expansion of the machinist occupation surpassed practically any other major group of craftsmen.

Despite the outstanding growth in their number, machinists did not experience a relative earnings decline. On the contrary, their relative earnings increased compared to most occupations from the early 1870s to the 1880s, and relative earnings gains seem to have disappeared only by the end of the century to some extent (see Table 3.1 and Section 3.5).[8] Taken together, the substantial employment expansion and relative earnings growth are consistent with a positive labor demand shock induced by the Second Industrial Revolution - relative to most other middle-skilled occupations.

3.3 Data

The main data sources for this work are various waves of the US full count census between 1850 and 1940 (Ruggles et al., 2021). This data set is complemented with i) novel, state- and time-varying earnings scores pre-1900 (Section 3.3.2); ii) newly digitized measures of county-level high school provision around 1880 (Appendix C.1.3); iii) the occupational education rank of Song et al. (2020); and iv) some development-related county characteristics from the NHGIS (Manson et al., 2021).

[8] One potential cause behind the disappearance of earnings gains as measured by occupational earnings scores is the following. While the machinist occupation was growing, it started to employ relatively more young, less experienced workers. Thus, a declining average experience level might have pushed the occupational earnings level down.

Table 3.1: Occupational earnings (1850-1892; in 1890 dollars)

	Yearly earnings score					Growth (%)			Growth (Massachusetts)	
Occupation	1850	1860	1870-72	1879-1881	1890-92	1850-1872	1872-1880	1872-1892	1872-1880	1872-1892
Blacksmith	453	462	541	427	523	19	-21	-3		1
Bricklayer		457	684	671	895		-2	31	-25	13
Cabinetmaker			400	430	487		7	22	-12	14
Carpenter	376	389	478	422	492	27	-12	3	-7	14
Locomotive engineer	568	542	654	758	874	15	16	34	6	49
Locomotive fireman	330	310	356	367	488	8	3	37	19	31
Machinist	**414**	**430**	**445**	**473**	**530**	**8**	**6**	**19**	**11**	**22**
Mason	398	459	580	535	734	46	-8	27	5	37
Painter	455	417	447	528	460	-2	18	3	3	12
Pattern maker	407	435	544	474	618	34	-13	14	3	44
Plasterer	429	414	613	625	766	43	2	25		
Shoemaker			456	380	454		-17	0		
Stone cutter		438	733	640	858		-13	17	-24	9
Teamster	364	290	344	369	447	-6	7	30	12	45
Watchman	269	270	290	288	362	8	-1	25	7	16

Note: the data source is U.S. Department of Labor (1900). Occupations are not harmonized. Earnings are converted to 1890 dollars using inflation values from `measuringworth.com`. Every yearly earnings score is constructed as follows. First, all state-year daily wage observations are collected which are based on at least ten individuals. For 1870-1872, 1879-1881 and 1890-1892, we take the state-year observation with the largest number of individuals. Second, the conversion of daily wage rates to yearly earnings is described in Appendix C.1.2. Finally, the values presented are the weighted averages of state-level scores. The weights are the number of individuals who contributed to the average wage calculation in every state. The last two columns contain only observations from Massachusetts.

3.3.1 Linking historical censuses

Analyzing intergenerational mobility necessitates linking individuals over time across distinct waves of the full count census. In this paper, we start out with the census conducted in 1870 to find the fathers (first generation - G1), whose offspring we follow in later decades and whose male parent (i.e. the grandfather - G0) we find in earlier decades in subsequent parts of this analysis.[9]

A few major restrictions are made on the 1870 full father (G1) sample. Exclusively fathers who were between 20 and 40 years old are included for two reasons. First, teenager workers tend to have transient occupations (Papageorgiou, 2014). Second, relatively old workers did not live with their kids anymore (the only way to identify family relationships) and were often not alive in 1900, the year chosen for the analysis of their long-run outcomes.[10] Furthermore, we exclude every individual with a non-occupational response or outlier wealth (personal property or real estate value above the 99$^{\text{th}}$ percentile). Individuals who held an agricultural occupation (farmer, farm manager/foreman/laborer), reported certain apprenticeship, or their harmonized occupation was a type of «not elsewhere classified» (e.g., 'Clerical and kindred workers (n.e.c.)') are also omitted. These restrictions are important because farmers had completely different characteristics compared to non-agricultural workers. Additionally, apprenticeships could obviously not be the final occupation of young adults. Finally, loosely classified occupations make the use of occupational education ranks or earnings scores less reliable if not

[9] The paper is limited to the analysis of male observations since the surname change of women upon marriage makes their linking over time impossible.

[10] The 1890 census records were burnt in a fire.

impossible.

As a next step, fathers are linked to their own 1900 observation. An individual is considered linked if at least one of the two conservative linking methods offered by Abramitzky et al. (2020) yields a match.[11] These linking methods have a particularly low false positive ratio (Bailey et al., 2020). Thus, we can avoid erroneously linking observations between two different people which helps us reduce the attenuation bias at the expense of a reduced sample size. Importantly, this linking rule is used for *every* linking in the entire paper.

In 1870, we can identify sons (G2) who lived with their father and link them to 1900 and 1940, separately. Exclusively sons who were at most 20 years old in 1870 are included. Then, we link these sons between 1870-1900 and 1870-1910, and find their kids in the respective end year in order to identify grandsons (G3). As a final step, we link grandsons found in 1900/1910 to 1940.

In Section 3.6.1 and 3.6.3, we use the characteristics of grandfathers (G0) in 1860. To do so, we link fathers back to 1850 and 1860. If a grandfather is only found in 1850 (e.g., because he already lived separately from the father in 1860), we link him forward to 1860 in order to obtain grandfathers' characteristics from the exact same year.

3.3.2 Occupational earnings scores for the late nineteenth century

One of our contributions is providing novel, state-specific earnings score estimates for the late-nineteenth-century United States. There are at least three reasons why these measures are crucial for this project. First, the traditional approach used in the literature - generating occupational income scores based on income reported in the 1940 census and using them in earlier decades - has been shown to perform more poorly the earlier it is applied prior to 1940 (Inwood et al., 2019; Saavedra and Twinam, 2020). Especially for periods when relative wages are changing rapidly, Inwood et al. (2019) recommend constructing earnings scores based on data from the studied time period, even if the sample might not be representative. Second, a considerable share of education received was informal in the 19th century (e.g., apprenticeships; see Goldin and Katz, 2008; Kelly et al., 2020; Meisenzahl and Mokyr, 2011). Therefore, while we can control for the (formal) education percentile rank devised by Song et al. (2020), we might not be able to capture the full difference in occupational human capital across occupations with this measure. However, earnings scores combined with the educational rank might very well capture the actual level of human capital implied by the sum of formal and informal

[11]The conservative linking methods provided by Abramitzky et al. (2020) require matches be unique by name and birthplace within a five-year age band.

education. Third, even the labor market of the north-eastern part of the United States (New England, Middle Atlantic, East-North Central), where most of the machinists lived, was not integrated until the 1880s and the difference between the north-eastern and Pacific (or southern) regions persisted even longer (Rosenbloom, 1996, 1998). Kaboski and Logan (2011) also find spatially-varying returns to education in the United States in the early twentieth century. Consequently, applying the same earnings score to a certain occupation all over the United States could lead to inaccurate conclusions. To the best of our knowledge, all existing earnings scores data sets for the late nineteenth century provide a single score for each occupation and pertain to the last decade of the 19th century (Preston and Haines, 1991; Sobek, 1996). Hence, we proceed to construct our own measure of state-specific occupational earnings for the 1870s and 1880s.

In this section, we outline the main steps of calculating these earnings scores. The interested reader can find detailed information and the discussion of the underlying assumptions in Appendix C.1.2. The source of our occupational earnings information is U.S. Department of Labor (1900). For many occupations,[12] we digitized the average daily wage found in 1870-72 (the 1872 score), 1879-81 (the 1880 score), 1890-92 (the 1892 score) in every state. In case of multiple observations within a three-year period, we digitized the daily wage which was calculated based on the largest number of observations. Then, daily wages were converted to yearly earnings scores and 1890 dollars. In this way, the earnings scores could be calculated for many large, low- and medium-skilled occupations. The income of high-skilled occupations (e.g., lawyers or physicians) was imputed by combining the earnings scores provided by Sobek (1996) with our own earnings scores.

The previously described steps provide *nominal* earnings scores. However, it is well-known that the costs of living differed significantly between urban and rural areas, and across states (Koffsky, 1949; Stecker, 1937). Hence, we also calculated *real* earnings scores adjusting for these price differences following Collins and Wanamaker (2014) (see Appendix C.1.2 for more details).

3.3.3 Summary statistics

Machinists were not the «representative agents» of the US economy. As it can clearly be seen from Table 3.2, most of their observables differed from the rest of the population. Machinist fathers in our analysis were slightly younger, more educated, less wealthy, more likely to be immigrants (especially of English ancestry) and lived in more urban, larger places than non-machinists in 1870. Since they were concentrated in the New England and Middle Atlantic census divisions,

[12] Besides machinists, the focus was on occupations i) which are in the control group in a large number in 1870 following propensity score matching, and ii) which played a large role in the economy later (i.e., important possible occupations for fathers or sons in 1900).

Table 3.2: Summary statistics of fathers (G1 in 1870)

Variable	Mean (*non-machinists*)	Difference (machinists - non-machinists) *Raw difference*	*Conditional on state-fixed effects*
Age (in years)	34,1	-0,5 [0.076] ***	-0,5 [0.089] ***
Literate (Yes=1)	0,92	0,05 [0.0082] ***	0,04 [0.0075] ***
Education rank of occupation (Song et al., 2020)	50,4	4,3 [1.182] ***	4,5 [0.968] ***
Value of real estates (in 1870 dollars)	793,4	-128,0 [32.218] ***	-74,9 [14.424] ***
Value of personal property (livestock, jewels, bonds, etc.)	350,9	-87,7 [15.703] ***	-85,9 [26.635] ***
Both parents native born (Yes=1)	0,58	-0,10 [0.025] ***	-0.10 [0.0229] ***
Both parents foreign born (Yes=1)	0,37	0,08 [0.0226]***	0,07 [0.0218] ***
Immigrant - UK or Ireland (Yes=1)	0,16	0,12 [0.0181] ***	0,1 [0.0173] ***
Immigrant - Germany (Yes=1)	0,15	-0,04 [0.0128] ***	-0,02 [0.009] *
Urban place of living (Yes=1)	0,44	0,34 [0.0195] ***	0,26 [0.0285] ***
Population of place of living	75532	33543 [14867] **	27692 [13249] **
New England (Yes=1)	0,13	0,16 [0.0731] **	-
Middle Atlantic (Yes=1)	0,32	0,04 [0.051]	-
East-north Central (Yes=1)	0,28	-0,1 [0.0354] **	-
West-north Central (Yes=1)	0,09	-0,04 [0.0220] *	-
South (Yes=1)	0,15	-0,05 [0.0249] **	-
West and Pacific (Yes=1)	0,03	-0,01 [0.0119]	-

Note: robust standard errors clustered at the state level (1870) in brackets. The summary statistics presented pertain to the final, total sample used in Table 3.5 and C6. The raw difference between means of machinists and non-machinists is the coefficient on the machinist dummy in an OLS regression with a constant and the dummy. This OLS regression also includes state-fixed effects (1870) in the last column. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

one might want to disentangle the effect of spatial distribution from other causes of significant difference. Therefore, differences in means are also presented after netting out state-fixed effects. Nonetheless, machinists seem to exhibit similar, though somewhat smaller differences in characteristics within states.

3.4 Empirical strategy

In this section, we describe our two, complementary empirical strategies: propensity score matching and fixed effects regressions.

3.4.1 Propensity score matching

Our primary empirical strategy is propensity score matching on many observable characteristics of fathers in 1870 (Austin, 2011; Ho et al., 2007; Leuven and Sianesi, 2003; Rosenbaum and Rubin, 1983). This estimation strategy amounts to estimating each individual's probability of being a machinist in a logit regression as a first step. For every machinist father, the five non-machinist fathers with the closest estimated probability are chosen as control observations with replacement.[13] Then, we compare the outcomes of machinist fathers and of their offspring to the outcomes of matched control fathers (and of their offspring) in the resulting sample. The relatively small share of machinists in the full sample implies that there are

[13] Additionally, we use a caliper of 0.01 and restrict the analysis to the common support of machinist and non-machinist fathers. This never results in losing more than ten treated observations in the main analysis. In a few analyses of later generations, we use ten instead of five neighbors because of the small sample size but this change is always duly noted.

many potential control observations, making our setting particularly well-suited for matching. The aim of matching is to reduce the correlation between the machinist dummy, which indicates if a father was a machinist in 1870, and the control variables. The full list of these control variables is shown in Appendix C.1.1. In short, we include i) personal characteristics (e.g., age, literacy, proxies of migration background, education rank of occupation, etc.); ii) place of living characteristics (e.g., urban dummy, state-fixed effects, measures of county-level industrialization, etc.); and iii) state-occupational level features constructed pre-1870 (e.g., probability of job switching or migration). Importantly, 1870 was the last historical census wave in which detailed information was collected on personal wealth: the value of real estates and personal property (the contemporary dollar value of all stocks, bonds, mortgages, notes, livestock, plate, jewels, and furniture owned by the respondent), separately.[14] Interactions and squares of many background characteristics are also included to match the distribution of these covariates more closely (Ho et al., 2007; Imai et al., 2008).

The main advantage of matching is that by reducing the correlation between the explanatory variable of interest and observables, such as personal wealth or urban status, we considerably reduce the influence of correlated unobservables. For example, the wealth proxies are most likely correlated with individual talent and family heritage, or the urban status can capture many urban (dis)amenities. Furthermore, matching diminishes our own discretion over how to control for a given background characteristic (Ho et al., 2007).[15]

The main limitation of using matching in our setting is that the full count census does not provide individual-level information on earnings and education before 1940. To overcome this lack of data, occupation-based characteristics are used. For education, the occupational education percentile rank of Song et al. (2020) is included. This is a percentile rank (0-100) based on the average occupational years of (formal) schooling in a person's birth cohort.[16] For income, which is probably more volatile over time than the education requirement of most occupations, we use our own state-level real earnings score constructed for 1870-72. The latter

[14]Wealth at a young age is an even better predictor of future wealth than parental wealth, and a good proxy for intergenerational correlation in savings behaviour and additional transfers from parents (Boserup et al., 2018).

[15]The application of propensity score matching in this paper is mostly immune to the criticism of King and Nielsen (2019) for several reasons: i) contrary to their claim that matching often increases imbalance compared to the unmatched sample, we transparently show that matching decreases it in our application; ii) the large sample makes the «propensity score matching paradox» less likely to appear; and iii) even though a caliper is used, the number of unmatched and, consequently, dropped machinists is always one-digit.

[16]For fathers and sons, we use the earliest available birth cohort around 1880 whose percentile rank is based on detailed years of schooling data and not merely on literacy. For grandsons, we use the percentile of the birth cohort around 1900.

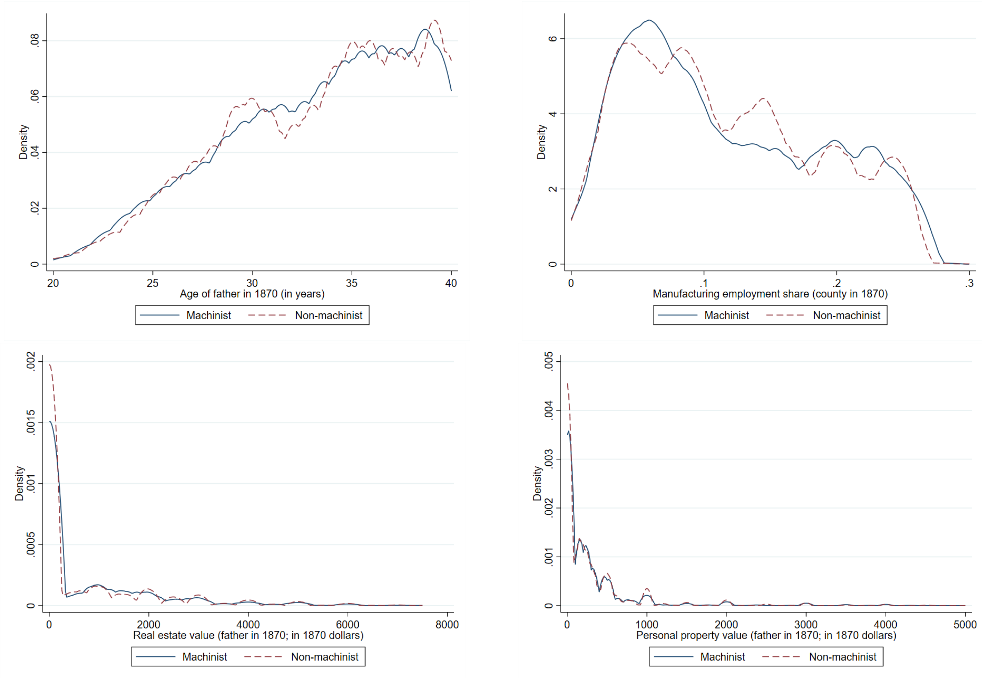


Figure 3.3: The histogram of some continuous characteristics after matching

Notes: the figures are created after the propensity score matching which generates Table 3.5. They depict the density of certain continuous control variables for machinists and matched non-machinists using weights obtained from the matching process.

is exclusively included in the analyses of income-related outcomes because its inclusion reduces the sample size along with the precision of the estimation without significantly changing the coefficients on non-pecuniary outcome variables. In fact, the real earnings score tends to be somewhat *lower* for machinists than for matched non-machinist control observations when it is not included in the list of control variables.[17]

In practice, propensity score matching works well in this setting and the correlation between observables and the machinist dummy, which is highly significant for most cases (Table 3.2), vanishes. Apart from similar means, the whole distribution of control covariates is closely matched (see Figure 3.3). However, the mean of a small subset of variables remains significantly different in some cases. The typical example is urban status: while machinists tend to be significantly *more*

[17]Machinists have an average score of $500, while the matched (unmatched) control average is $530 ($582) in Table 3.5. Notice that this imbalance works *against* our findings.

urban compared to the full sample, they are somewhat *less* urban in the matched one. Nevertheless, the standardized difference lies below 10% (the upper bar for tolerable difference - Austin, 2011) even in this case.[18] To avoid any bias from such residual differences, we include every control variable (their main effects) which has a significantly different mean at 5% in a regression after matching.[19] The other reason for running a regression on the matched sample instead of reporting the immediate outcome of matching is to construct clustered, more conservative standard errors at the state level.

The occupations with the most matched control observations are presented in Table C1. While the role of carpenters, truck and tractor drivers, and shoemakers is relatively large, none of them exceeds 10% of the control observations. We also show in Section 3.6 that their omission does not affect the results in any meaningful way. It must be emphasized that we use harmonized occupational codes provided by IPUMS as it is usually done in the literature. Therefore, the category 'Truck and tractor drivers' mainly consists of teamsters, draymen and hackmen in the 19$^{\text{th}}$ century.

3.4.2 Fixed effects regression

Despite the appealing features of propensity score matching, it precludes the inclusion of numerous fixed effects for two reasons: the algorithm occasionally does not converge when including county or county-urban status-fixed effects, and the small size of the matched subsample makes the estimation of fixed effects very imprecise. Another problem with matching is that it does not allow for weighting, so the sample cannot be weighted to make it representative of the US population (more details in Section 3.6.2). To address these issues, we also present some results using fixed effects regressions.

In our fixed effects regressions, exactly the same baseline controls are included as in matching in addition to county-fixed effects (1870).[20] Therefore, the offspring of machinists are compared to the offspring of non-machinists who lived in the exact same county in 1870 and had similar paternal (G1) observables. To bring this analysis in spirit closer to matching, fathers whose occupation is below the 25$^{\text{th}}$ or above the 85$^{\text{th}}$ educational rank percentile are omitted from the analysis (the rank

[18]In Table 3.3, the difference (machinist minus non-machinist) between the probability of urban place of living in 1870 is 34% before matching and -4% after matching. In this particular application, the mean (median) standardized bias is 15.4 (9.2) before matching and 2.2 (1.3) after matching.

[19]We are aware of the «balance test fallacy» coined by Ho et al. (2007) and Imai et al. (2008), who discourage researchers to use the significance of difference between means as a balancing threshold. However, we find in practice that the inclusion of significantly different (p-value below five percent) characteristics matters to a very limited extent and the inclusion of non-significantly different variables does not have any effect on the estimation.

[20]We use the *reghdfe* package in Stata by Correia (2016).

of machinists is the 55th). In this way, the very low-skilled (e.g., lumbermen or miners) and high-skilled (e.g., architects or lawyers) fathers are not in the sample so that we can focus on the «middle class». Another advantage of fixed effects regressions is that they allow us to precisely estimate interaction terms between the machinist dummy and other variables as well.

Formally, the regression specification takes the following form:

$$y_{s,f,c,1900} = \beta \cdot Machinist_{f,1870} + \gamma \cdot x'_{f,1870} + \delta_{c,1870} + \epsilon_{s,f,c,1900} \tag{3.1}$$

where $y_{s,f,c,1900}$ represents an outcome variable for son s of father f measured in 1900 (e.g., a binary variable if the son held an agricultural occupation). The explanatory variable of interest is $Machinist_{f,1870}$, which equals one if the father was a machinist in 1870. County-fixed effects ($\delta_{c,1870}$) and all paternal baseline controls ($x_{f,1870}$) are also included. Reassuringly, the effects on main outcomes estimated by propensity score matching and fixed effects regressions tend to be quantitatively and qualitatively very similar.

In order to get a consistent estimate of β, the error term, $\epsilon_{s,f,c,1900}$, must be uncorrelated with the machinist dummy conditional on our predetermined controls. Thus, the main concern about the validity of the empirical strategy is that particularly talented fathers sorted into the machinist occupation before 1870 in an unobserved way, causing omitted variable bias. To alleviate this concern, we present many heterogeneity and robustness checks in Sections 3.5.2 and 3.6. These empirical exercises suggest that (a within-family intergenerational transmission of) unobserved ability is not driving our results.

3.5 Main results

In the first part of this section, key results establishing the gains of the machinist occupation post-1870 and the intergenerational transmission between machinists and their (grand)sons are presented. To elaborate on mechanisms of transmission, we conduct some heterogeneity exercises in the second part.

3.5.1 Long-term effects and intergenerational transmission

Fathers (G1) between 1870 and 1900 Table 3.3 contains the main, non-pecuniary outcomes for our linked 1870-1900 father sample using propensity score matching. The first column shows that machinists were 8.7 percentage points (0.2 standard deviation) less likely to switch their occupation. This coefficient can be decomposed into switching to different types of jobs. In particular, roughly one-third of the total effect stemmed from a lower likelihood of switching to an agricultural job (Column 2), while the rest can be attributed to a less likely change for another

Table 3.3: Main outcomes - fathers (G1; 1870-1900)

	Occupational change [(1) = (2)+(3)]			Migration (Yes=1)		Place of living (1900 - Yes=1)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Any occupation (Yes=1)	Agricultural	Non-agricultural	Within-state	Across states	Higher population than in 1870	Urban
Machinist (G1)	-0.087***	-0.033***	-0.054***	0.006	0.011	0.058***	0.073***
	(0.010)	(0.006)	(0.012)	(0.007)	(0.008)	(0.012)	(0.010)
Mean of outcome	0.77	0.19	0.57	0.20	0.37	0.46	0.50
Standard deviation of outcome	0.42	0.39	0.49	0.40	0.48	0.50	0.50
Unbalanced controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	18811	18811	18811	18811	18811	18811	18811
Number of clusters	50	50	50	50	50	50	50

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1900) level. All specifications are weighted by weights obtained from propensity score matching described in the main text. The summary statistics reported are unweighted and pertain to the full estimation sample before matching. The final sample includes 3902 matched machinist fathers. The outcome variable is a binary variable which equals one if the father changed occupation (Col. 1), changed occupation and the new occupation is agricultural (Col. 2 - farmer, farm manager/foreman/laborer) or non-agricultural (Col. 3), migrated within-state across counties (Col. 4) or across states (Col. 5), his place of residence fell into a larger *SIZEPL* category in 1900 than in 1870 (Col. 6), he lived in an urban place in 1900 (Col. 7). Unbalanced controls included in the regressions are characteristics whose mean between machinist and control fathers is still significantly different at 5% after matching. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

non-agricultural occupation (Column 3). We interpret the lower likelihood of leaving the initial occupation as the first sign of a beneficial effect on machinists post-1870. Namely, there is an extensive literature which documents the large costs of occupation switching in many contexts (e.g., Artuç et al., 2010; Cortes and Gallipoli, 2018; Dix-Carneiro, 2014; Kambourov and Manovskii, 2009; Sanders and Taber, 2012; Traiberman, 2019). This literature suggests that machinists lost less lifetime earnings caused by the costly accumulation of occupation- or task-specific human capital due to their lower likelihood of changing their occupation.

Internal migration has been established as a pre-eminent way to upward mobility in the studied time period (Long and Ferrie, 2007, 2013; Ward, forthcoming). However, no evidence is found on a differential probability of migration within or across states (Columns 4-5). We further elaborate on migration *destinations* in Table C2. First, we decompose the insignificant migration differential and find that machinist fathers tended to migrate significantly more (less) to urban (rural) places. Second, we also establish that initially rural machinists were particularly more likely to move to urban areas and initially urban machinists were less likely to migrate to rural areas. These effects can clearly be seen in Columns 6-7 of Table 3.3 as well: machinist fathers lived in more populous and more urban places by 1900 (both effects stronger than 0.1 standard deviation). The urban environment could provide them and their offspring with better opportunities in a period when urbanization and growth were tightly intertwined.

Next, we direct our attention to analyze the effect on occupational earnings scores. In Columns 1-2 of Table 3.4, we assume that fathers held the same occupation and lived in the same place in 1880 as in 1870. We do so because an additional linking to 1880 would come at the expense of a large sample size reduction. The coefficients suggest that machinist fathers experienced a relative increase of 8-9 log-points in their earnings score.[21] This finding is unsurprising since it is documented in Table 3.1 that the relative wage of the machinist occupation increased compared to most other occupations in this time period. While the magnitude of the effect is substantial (0.25 s.d.), we treat it as an *upper* bound on the actual effect because control fathers could switch their occupation or place of living in order to reduce the relative earnings gap. Therefore, we also linked fathers between 1870 and 1880 (instead of 1900) and, thus, allowed for occupation and place of living change in Table C3. As expected, the estimated earnings effect is somewhat smaller (6-7 log-points) but still significantly positive.

In the last four columns of Table 3.4, we use the occupation and state of living of fathers in 1900 to construct outcome variables. The first observation is that both the nominal and real earnings score gains expectedly declined compared to 1880. The second observation is that using the widely-used earnings scores (Preston and Haines, 1991; Sobek, 1996) results in a larger coefficient compared to our own nominal score.[22] We suspect that this discrepancy partly stems from the treatment of agricultural workers. In particular, the ratio between the score of farm laborers and other laborers is substantially lower in Sobek (1996) or Preston and Haines (1991) than in the case of our scores. Knowing that machinists were significantly less likely to switch to agricultural occupations, assigning lower scores to agricultural jobs amplifies the relative earnings gains of machinists. We believe that our scores might be more accurate since Alston and Hatton (1991) or Hatton and Williamson (1991) show that a large part of the gap in nominal earnings between farm and common laborers can be explained by more in-kind benefits (especially the value of accommodation) for the former group. As explained in Appendix C.1.2, we calculate farm laborers' remuneration based on daily wages *without* accommodation which brings the ratio between the earnings of farm and common laborers close to those reported in Alston and Hatton (1991) and Hatton and Williamson (1991), and takes into account the monetary value of accommodation. Nevertheless, the 3.5-8 log-points higher nominal earnings scores do not account for the fact that machinist fathers were more likely to reside in more populous, urban places in 1900 - implying higher consumer prices. When these differences in cost of living are adjusted for, the estimated positive effect becomes insignificant (Column 5). In other words, the real gains of initially machinist fathers were arbitraged away in the (very) long run.

We further investigate the effect on earnings scores in Table C4, focusing on

[21] The same coefficient on the nominal and real earnings score is mechanical. Since we assume that fathers do not change their occupation, state and urban status between 1870 and 1880, only the nominal wage change of the given occupation matters in this calculation.

[22] Preston and Haines (1991) do not provide an earnings score for owner-occupier farmers and calculate earnings scores based on an urban sample in the Cost of Living survey. Sobek (1996) instead calculates an unweighted average of all distinct earnings scores for every occupation.

Table 3.4: Measures of economic status (medium- and long-run) - fathers (G1)

	State-occupation in 1870		State-occupation in 1900			
	(1) State-level nominal log-score (1880)	(2) State-level real log-score (1880)	(3) Sobek log-score	(4) State-level nominal log-score (1892)	(5) State-level real log-score (1892)	(6) Preston-Haines log-score
Machinist (G1)	0.085*	0.085*	0.067***	0.034***	0.019	0.084***
	(0.048)	(0.048)	(0.009)	(0.011)	(0.012)	(0.012)
Mean of outcome	6.11	6.25	6.21	6.25	6.37	6.49
Standard deviation of outcome	0.33	0.34	0.56	0.44	0.43	0.37
Unbalanced controls	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	16124	16124	11573	11573	11573	9895
Number of clusters	32	32	47	47	47	50

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1870 in Col. 1-2; 1900 in Col. 3-6) level. All specifications are weighted by weights obtained from propensity score matching described in the main text. The summary statistics reported are unweighted and pertain to the full estimation sample before matching. The final sample includes 3820 (Col. 1-2), 2669 (Col. 3-5) and 2334 (Col. 6) matched machinist fathers. The outcome variable is the state-level nominal and real log-score (Col. 1 and 2.) merged to the state and occupation of fathers in 1870; the Sobek, state-level nominal and real, and Preston-Haines log-score (Col. 3-6) merged to the state and occupation of fathers in 1900. Unbalanced controls included in the regressions are characteristics whose mean between machinist and control fathers is still significantly different at 5% after matching. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

individuals who changed their occupation between 1870 and 1900. The main takeaway of this table is that, besides the increasing relative wage of the machinist occupation, the relative earnings gain of initially machinist fathers was the result of a six percentage points lower likelihood of switching to an occupation with considerably lower earnings rather than differential upward mobility. In our interpretation, non-machinists lost their occupations more frequently in the turbulent times of the Gilded Age, when recurrent busts in the aftermath of panics characterized an overall robust growth. As breadwinners of their family, they had to find an alternative, potentially lower-paying (agricultural) occupation in the absence of generous unemployment benefits. This interpretation is consistent with Boone and Wilse-Samson (2019) who show that movement to farms served as a source of migratory insurance during the Great Depression. Moreover, the fact that machinists were not more likely to switch to managerial jobs or becoming proprietors (Column 6) supports our claim that the improved outcomes for themselves and their offspring were not the result of unobserved talent.

The outcomes of sons (G2) The next question we answer is if the benefits of fathers could be transmitted to their sons. The main, non-pecuniary outcomes are presented in Table 3.5. Similarly to their fathers, sons were significantly less likely to hold an agricultural occupation (Column 1). Furthermore, they held occupations which had significantly higher education ranks (almost +0.1 s.d.). Whereas the latter finding simply suggests that machinists' sons held occupations with on average more educated peers, we can estimate *individual-level* schooling using the 1940 census. We linked sons between 1870 and 1940 to this end. The results in Table C5 show that machinists' sons had indeed 0.21 years more schooling. The effect is mainly the result of a 3.6 percentage points (+0.1 s.d.) higher likelihood of having some secondary education, meanwhile the effect on university education is a tightly estimated zero. The secondary school coefficient should be treated as a *lower* bound on the actual effect since the beneficial effect of education on longevity could lead to endogenous attrition. Thus, as sons were at least seventy years old in 1940, the less educated control sons might have been more likely to pass away before 1940, leading to a downward bias in the estimated coefficient.

In line with the higher level of educational attainment, we find a significantly higher probability of long-distance migration for sons between 1870 and 1900 (+0.1 s.d. - Column 4 in Table 3.5; Malamud and Wozniak, 2012; Rosenbloom and Sundstrom, 2003; Wozniak, 2010). This foreshadows our findings on higher earnings because the migration premium increased in distance in this time period (Ward, forthcoming). The effect on a higher probability of urban homes and larger cities persists, though it slowly starts to fade away compared to the first generation.[23] In fact, the difference in the manufacturing employment share of the county of living in 1900 is insignificant. This suggests a certain convergence in the type of place of living across the sons of machinists and non-machinists. Finally, we uncover some evidence that the more educated sons of machinists had fewer kids, perhaps because they faced higher opportunity costs of raising children (Ager et al., 2020b). The effect on marriage probability and house ownership is insignificant.

The pattern of the earnings effect for sons is similar to the paternal one: the well-known nominal scores having a more positive coefficient than our own score, and a diminished coefficient once across-state and rural-urban price differences are accounted for (Table 3.6).

The outcomes of grandsons (G3) Table 3.7 documents the main, non-pecuniary outcomes for machinists' grandsons. The set of possible outcomes is richer thanks to the increased data collection effort in the 1940 census. First, we learn that even the grandsons were less likely to be engaged in an agricultural occupation, they worked more weeks, but did not have a differential likelihood of self-employment (Columns 1-3). The availability of individual-level educational attainment allows us to compare the magnitude of the effect on the occupational education rank and on the highest grade of individual-level schooling (Columns 4 and 5). Reassuringly, both variables imply a very similar, positive magnitude: 0.1 standard deviation. This comparison corroborates our entire analysis because we seem to approximate actual education very closely with occupation-level average education scores. We also see that machinist grandsons were almost five percentage points more likely to have completed at least primary school. However, we do not find any significant effect on the number of children, marriage probability and house ownership

[23]For instance, the positive effect of an urban place of living drops by 60% in magnitude, from 1.5 to 0.9 standard deviations.

Table 3.5: Main outcomes - sons (G2; 1900)

	Occupational characteristics		Migration (Yes=1)		Place of living			Personal characteristics		
	(1) Agricultural occ. (Yes=1)	(2) Education rank	(3) Within-state	(4) Across states	(5) Urban (Yes=1)	(6) Higher population than in 1870 (Yes=1)	(7) Manuf. emp. per capita (county)	(8) # of children	(9) Married (Yes=1)	(10) Owning house (Yes=1)
Machinist (G1)	-0.028***	2.403***	-0.009	0.043***	0.045***	0.033**	0.004	-0.080**	0.005	0.000
	(0.007)	(0.745)	(0.011)	(0.012)	(0.011)	(0.016)	(0.003)	(0.038)	(0.010)	(0.014)
Mean of outcome	0.19	49.68	0.30	0.31	0.55	0.52	0.09	1.70	0.77	0.43
Standard deviation of outcome	0.39	28.25	0.46	0.46	0.50	0.50	0.07	1.86	0.42	0.50
Unbalanced controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	8745	8745	8745	8745	8745	8745	8745	8745	8745	8745
Number of clusters	45	45	45	45	45	45	45	45	45	45

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1900) level. All specifications are weighted by weights obtained from propensity score matching described in the main text. The summary statistics reported are unweighted and pertain to the full estimation sample before matching. The final sample includes 1842 matched machinist sons. The outcome variable is an agricultural occupation indicator (Col. 1 - farmer, farm manager/foreman/laborer), the education rank of occupation (Col. 2), a binary variable which equals one if the son migrated within-state across counties (Col. 3) or across states (Col. 4) between 1870 and 1900, a binary variable which equals one if the son lived in an urban place in 1900 (Col. 5) or his place of residence fell into a larger *SIZEPL* category in 1900 than in 1870 (Col. 6), manufacturing employment per capita (as % of total county population; Col. 7), the number of children in the household (Col. 8), a marriage status (Col. 9) and house ownership (Col. 10) indicator. Unbalanced controls included in the regressions are characteristics whose mean between machinist and control fathers is still significantly different at 5% after matching. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6: Measures of economic status - sons (G2; 1900)

	(1) Sobek log-score	(2) State-level nominal log-score	(3) State-level real log-score	(4) State-level real score (level)	(5) Preston-Haines log-score
Machinist (G1)	0.070***	0.042***	0.032***	15.208**	0.069***
	(0.014)	(0.011)	(0.010)	(6.452)	(0.011)
Mean of outcome	6.23	6.23	6.35	628.89	6.45
Standard deviation of outcome	0.53	0.43	0.41	294.05	0.39
Unbalanced controls	Yes	Yes	Yes	Yes	Yes
Sample size	6812	6812	6812	6812	6687
Number of clusters	45	45	45	45	45

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1900) level. All specifications are weighted by weights obtained from propensity score matching described in the main text. The summary statistics reported are unweighted and pertain to the full estimation sample before matching. The final sample includes 1548 (1514 in Col. 5) matched machinist sons. The outcome variable is the Sobek, state-level nominal and real log-score (Col. 1-3), the state-level real score in levels (Col. 4) and the Preston-Haines log-score (Col. 5). Unbalanced controls included in the regressions are characteristics whose mean between machinist and control fathers is still significantly different at 5% after matching. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

indicator (though the signs of the coefficients are in the expected direction).

The final piece of main results concerns the income of grandsons (Table 3.8). The first four columns include wage earner as well as self-employed grandsons. As self-employed individuals did not report their income, we impute it following the best practice in the literature (see Appendix C.1.6). The results are clear: both the nominal and the real wage effect are positive and significant. This conclusion becomes even stronger when we focus exclusively on wage earners whose wages do not require imputation (Columns 5-8).

In conclusion, we document large and significant gains for the sons and grandsons of machinists in terms of education- and income-related outcomes even after seventy years. The implied limited level of intergenerational mobility[24] is consistent with a large literature which demonstrates that intergenerational mobility was indeed low and declined at the turn of the twentieth century in the United States (Long and Ferrie, 2013; Olivetti and Paserman, 2015; Song et al., 2020; Ward, 2019). Therefore, the initial gains of machinist fathers dissipated slowly over time and generations.

[24]Around 65% of the earnings gains of fathers (Column 1 in Table C3) were transmitted to their sons (Column 2 in Table 3.6), which is consistent with the values reported in Ward (2019).

Table 3.7: Main outcomes - grandsons (G3; 1940)

	Occupational characteristics				Education		Personal characteristics		
	(1) Agricultural occ. (Yes=1)	(2) Weeks worked	(3) Self-employed (Yes=1)	(4) Education rank	(5) # of grades completed	(6) More than primary education (Yes=1)	(7) # of children	(8) Married (Yes=1)	(9) Owned a house (Yes=1)
Machinist (G1)	-0.024**	0.865**	0.013	1.902***	0.312***	0.047**	-0.002	0.016	0.023
	(0.011)	(0.408)	(0.014)	(0.666)	(0.113)	(0.020)	(0.061)	(0.012)	(0.022)
Mean of outcome	0.10	44.29	0.22	31.95	9.90	0.54	1.52	0.87	0.53
Standard deviation of outcome	0.30	14.33	0.42	19.04	3.26	0.50	1.66	0.33	0.50
Unbalanced controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	6383	6383	6383	6383	6383	6383	6383	6383	6383
Number of clusters	49	49	49	49	49	49	49	49	49

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1940) level. All specifications are weighted by weights obtained from propensity score matching described in the main text. The summary statistics reported are unweighted and pertain to the full estimation sample before matching. The final sample includes 969 matched machinist grandsons. We use ten matched control observations instead of five owing to the small number of machinist grandsons. The outcome variable is an agricultural occupation indicator (Col. 1 - farmer, farm manager/foreman/laborer), the number of weeks worked (Col. 2), a self-employed status indicator (Col. 3), education rank of occupation (Col. 4), highest grade of schooling (Col. 5 - winsorized at the 99th percentile in the final sample), a binary variable which equals one if the highest grade of schooling is at least 9 years (Col. 6), number of children in the household (Col. 7), a marriage status (Col. 8) and house ownership (Col. 9) indicator. Unbalanced controls included in the regressions are characteristics whose mean between machinist and control fathers is still significantly different at 5% after matching. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.8: Measures of income - grandsons (G3; 1940)

	Self-employed & wage workers				Wage workers			
	(1) Log-wage (nominal)	(2) Log-wage (real)	(3) Wage (level)	(4) Non-wage income (Yes=1)	(5) Log-wage (nominal)	(6) Log-wage (real)	(7) Wage (level)	(8) Non-wage income (Yes=1)
Machinist (G1)	0.068**	0.061*	112.774**	-0.009	0.082**	0.074**	117.019**	-0.026**
	(0.034)	(0.034)	(52.069)	(0.012)	(0.036)	(0.037)	(55.541)	(0.011)
Mean of outcome	7.26	0.20	1855.92	0.30	7.21	0.15	1771.97	0.17
Standard deviation of outcome	0.82	0.81	1278.30	0.46	0.82	0.81	1207.89	0.38
Unbalanced controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	6212	6212	6212	6212	5244	5244	5244	5244
Number of clusters	49	49	49	49	49	49	49	49

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1940) level. All specifications are weighted by weights obtained from propensity score matching described in the main text. The summary statistics reported are unweighted and pertain to the full estimation sample before matching. The final sample includes 908 (746 in Col. 5-8) matched machinist grandsons. We use ten matched control observations instead of five owing to the small number of machinist grandsons. The outcome variable is the log of reported nominal wage (Col. 1 and 5 - winsorized at the 95th percentile in the final sample), the log of reported real wage (Col. 2 and 6 - winsorized at the 95th percentile in the final sample), the level of reported nominal wage (Col. 3 and 7 - winsorized at the 95th percentile in the final sample), and a meaningful non-wage income indicator (Col. 4 and 8. - more than $50). The sample includes wage workers as well as self-employed people reporting non-zero wage in Columns 1-4, while it is restricted to wage earners in Columns 5-8. The imputation of self-employed income is described in Appendix C.1.6. Unbalanced controls included in the regressions are characteristics whose mean between machinist and control fathers is still significantly different at 5% after matching. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.5.2 Mechanisms behind the intergenerational transmission

The goal of this section is to understand the mechanism behind the intergenerational transmission of the improved socio-economic status of machinist fathers to their offspring. We focus on the transmission from fathers to sons since the small sample size for grandsons does not let us draw robust conclusions.

Secondary education as a pathway to upward mobility Education meant at most primary schooling for the overwhelming majority of young people in the late-nineteenth-century United States: merely nine percent of American youth had high school diploma even in 1910. This share only moderately increased from the 1870s until the start of the so-called High School Movement in the 1900s. In the studied time period, high schools were mostly attended by the children of the (upper)-middle class. To a lesser extent, farmers or manual workers also sent their offspring to study as they saw high school education as a way out of a rural life and physical toil for their children. Rural areas maintained mostly private high schools and only cities could afford to finance public high schools. Private secondary schools regularly charged a tuition fee and non-residents were expected to pay a boarding fee (cost of accommodation) as well, meanwhile public institutions normally did not demand any payment. Nevertheless, the role of public schools remained inferior to private institutions until the 1890s. Therefore, in the absence of strictly implemented compulsory schooling laws for secondary schooling, it mainly depended on their parents' income and preferences if the sons of machinists and their peers received post-primary education (e.g., Goldin, 1998; Goldin and Katz, 2000, 2008; Lingwall, 2010; Tyack, 1974).

We documented earlier that machinist fathers experienced occupational stability and higher earnings in the period when most sons in the sample reached high school age around 1880. Thus, they could afford to educate their sons more easily. Indeed, we present evidence consistent with a complementarity between local private secondary school provision and parental income. Additionally, it has been demonstrated that parents of similar socio-economic status tend to have similar preferences over the schooling of their kids (Boneva and Rauh, 2018). Thus, we do not expect these (unobserved) preferences to drive the findings. Moreover, our subsequent findings are inconsistent with preferences for more schooling at every tuition fee (cost of education) level.[25]

First, the effect of private high school provision is studied by interacting the machinist main effect with the share of boys who attended high school in the the county. We assume that sons still lived in the county where they were located in 1870 when they reached high school age. In addition, only those sons are included who were not older than ten years in 1870, so that they were not too old to benefit from secondary education and reached high school age around 1880 - the year which our schooling measure corresponds to. Every specification includes an interaction with the county-level share of manufacturing employment as well, so that we can avoid that the results are driven by the known negative association between high schooling and industrialization (Goldin and Katz, 1999). Both the high school provision and industrialization proxy are standardized in the full sample. This means that the coefficients can be interpreted as the effect of one standard deviation increase in the given variable.

The results are presented in Table 3.9. When the tuition fee was neither too cheap (so the main cost of schooling was the foregone wage and practically everyone could attend high school; Column 2) nor prohibitively expensive even for machinists (Column 4), the sons of machinists benefited from the increased availability of private high schools. At mean private high school provision, a machinist son had an occupation with a four percentiles higher education rank. If he instead grew up in a county with a one standard deviation lower high school provision, the entire positive effect might have vanished (Column 3). To strengthen our increased parental investment interpretation, we show that the identified positive coefficient on the interaction term is driven by counties which had low public high school provision, i.e. boys could mainly pursue secondary education at private schools as public high school provision was very limited (Column 5). Additionally, Column 6 establishes that the coefficient on the interaction term is particularly large across counties with high-quality private secondary schools (high teacher-pupil ratio - Card and Krueger, 1992; Chetty et al., 2014). The last two columns show that the complementarity between a machinist father (income effect) and local private high school provision also manifests itself for other relevant outcomes.

[25]The entire schooling data collection and preparation process is described in Appendix C.1.3.

Table 3.9: Heterogeneity by the level of private tuition fee - sons (G2; 1900)

	Education rank						Other outcomes (medium fee)	
	(1) Full sample	(2) Low tuition fee (<25th percentile)	(3) Medium tuition fee	(4) High tuition fee (>75th percentile)	(5) Medium tuition fee (below median public HS)	(6) Medium tuition fee (teacher-pupil ratio>0.04)	(7) Max. primary education (% in occupation)	(8) Sobek log-score
Machinist (G1)	4.207***	6.454***	4.000***	3.143	4.369**	1.477	-2.067**	0.091***
	(0.928)	(1.491)	(1.270)	(1.900)	(2.074)	(1.625)	(0.869)	(0.024)
Private high school (%) x Machinist (G1)	0.893*	-2.225	3.574***	0.292	4.803***	4.368***	-2.145***	0.035*
	(0.512)	(1.543)	(0.973)	(0.854)	(1.278)	(1.191)	(0.794)	(0.020)
Manufacturing emp. (%) x Machinist (G1)	0.058	0.094	0.526	-0.131	0.510	1.979**	-0.390	-0.016
	(0.491)	(0.978)	(0.637)	(1.124)	(1.088)	(0.926)	(0.477)	(0.013)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County-fixed effects (1870)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	39570	9871	19742	9957	6455	13511	19742	19181
Number of clusters	45	45	45	45	45	45	45	45

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1900) level. None of the specifications is weighted. The sample includes all sons who were not older than ten years in 1870, and whose father held an occupation between the 24.7th and 84.7th education rank percentiles in 1870. The sample is additionally restricted to sons who lived in a county in 1870 i) with private tuition fee below the 25th percentile (Col. 2) / between the 25th and 75th percentiles (Col. 3 and 5-8) / above the 75th percentile (Col. 4); ii) with below median public high school share (male public high school students as % of 14-20 year-old males in the county in 1880 - Col. 5); iii) with an average teacher-pupil ratio above 0.04 in private schools. The outcome variable is the education rank of occupation (Col. 1-6), the share of workers who had at most primary education in the son's occupation (Col. 7) and the Sobek log-score (Col. 8). The share of private high school students and of manufacturing employment (as % of county population in 1870) are winsorized at the 99th percentile and standardized. Baseline controls are described in Appendix C.1.1. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Better private high school provision at medium tuition fee level led to machinists' sons having fewer people with at most primary education in their occupation (Song et al., 2020) and a larger increase in their earnings score.

Second, we study the effect of public secondary education supply in large cities (at least 7.500 inhabitants in 1880). In the standard model of Becker and Tomes (1986), the effect of a higher income of machinist fathers allowing their sons to stay in school longer should be diminishing in the expansion of the mostly free of charge, public high school system. This happens if parents faced restrictions to borrowing or savings and public schooling purely substituted for private schooling.[26] The test of this hypothesis is presented in Table 3.10. At mean public high school provision, an urban machinist's son had a three percentage points higher occupational education rank compared to sons of non-machinists. However, half of this relative gain was lost in counties with one standard deviation higher public high school provision (e.g., Akron, OH, Hartford, CT or Richmond, VA). Compared to these places, the gains of machinists' sons were three times larger in cities with one standard deviation below the mean (e.g., Indianapolis, IN, Jersey City, NJ or Joliet, IL). Columns 2 and 3 show that the other two outcomes of interest were influenced by expanding public secondary schools in a similar way. Column 4 establishes that the expansion of public schools particularly mattered under medium private tuition fee, in line with the previous analysis of private high schools. Exclusively urban sons were included in the estimation so far, even though Goldin and Katz (2008) write that township public schools sometimes educated the youth of the urban center as well as those of nearby rural communities. Therefore, the sample is expanded with rural sons within the county of large cities in Column 5. The interaction coefficient becomes somewhat smaller, suggesting that the effect is driven by the urban subsample who grew up in the physical proximity of schools. In Column 6, cities with more than 100.000 inhabitants are excluded from the sample which makes the interaction term even larger in magnitude.

The extent of local public high schooling was influenced by other factors than the level of industrialization (high opportunity cost of staying in school in industrialized counties) as well. Wealthier, more equal and stable communities tended to be associated with a more abundant public high school supply. Using proxies following Goldin and Katz (1999), we demonstrate that our interaction with public schooling does not capture, for instance, the beneficial effect of wealthier residents who, in turn, were willing to invest in public schools. In Column 7, two wealth proxies are included, but the coefficient of interest remains unaffected. We use the wealth share of the top 1% of residents as a proxy for wealth inequality and the

[26]Goldin and Katz (2008) report that the tuition fee itself was on average 5% of the gross earnings of skilled workers. The boarding fee could double or triple the costs. The recent empirical evidence on credit constraints is discussed in the introduction.

Table 3.10: Heterogeneity by the supply of public schooling - sons (G2; 1900)

	Full urban sample			Dependent variable: education rank				
	(1) Education rank	(2) Max. primary educ. (% in occ.)	(3) Sobek log-score	(4) Medium tuition fee	(5) Rural and urban county population	(6) No large cities	(7) Wealth proxies	(8) Inequality and old population
Machinist (G1)	3.033***	-1.770***	0.029	4.557***	3.179***	2.724*	2.804**	1.991*
	(1.068)	(0.610)	(0.018)	(1.640)	(0.793)	(1.396)	(1.168)	(1.060)
Public high school (%) x	-1.644***	1.031***	-0.022*	-2.587***	-1.302**	-2.379***	-1.475***	-1.940***
Machinist (G1)	(0.588)	(0.359)	(0.012)	(0.743)	(0.522)	(0.769)	(0.540)	(0.526)
Manufacturing emp. (%) x	0.534	-0.183	0.011	-0.296	0.703	0.438	0.819	-0.822
Machinist (G1)	(0.559)	(0.303)	(0.012)	(0.783)	(0.511)	(0.826)	(0.674)	(0.972)
Agricultural production per							2.234	
agric. worker x Machinist (G1)							(1.496)	
Wealth per capita x							-0.613	
Machinist (G1)							(1.061)	
Top 1% share of wealth x								4.781**
Machinist (G1)								(1.863)
Elderly population								2.239*
(%; above 65 y.o.) x Machinist (G1)								(1.317)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County-fixed effects (1870)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	19393	19393	18767	9901	26555	12424	19393	19393
Number of clusters	45	45	45	45	45	45	45	45

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1900) level. None of the specifications is weighted. The sample includes all sons who were not older than ten years in 1870, whose father held an occupation between the 24.7th and 84.7th education rank percentiles in 1870, and who lived in places classified as urban in 1870. Additionally, the sample is restricted to sons in counties with medium tuition fee (Column 4 - see Table 3.9), is expanded to include rural sons as well within a county (Column 5), and does not include sons in cities with more than 100.000 inhabitants in 1870 (Column 6). The outcome variable is the education rank of occupation (Col. 1 and 4-8), the share of workers who had at most primary education in the son's occupation (Col. 2) and the Sobek log-score (Col. 3). The share of public high school students (as % of 14-20 years old males in the county in 1880), the share of manufacturing employment (as % of county population in 1870), the agricultural production per agricultural worker (the estimated yearly, county-level agricultural production is from Manson et al. (2021), while the number of agricultural workers is from the 1870 full count census - agricultural workers are farmers, farm managers, foremen and laborers), the wealth per capita (calculated as the total wealth in a county - the sum of real estates and personal property - divided by county population in 1870), the top 1% share of wealth (calculated as the county-level wealth - sum of real estate and personal property - share of the richest one percent in 1870; only males who were above 16 years old), and share of elderly people (share of people older than 65 years in the 1870 full count census) are winsorized at the 99th percentile and standardized. Baseline controls are described in Appendix C.1.1. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

share of elderly people to capture the stability of the local community in Column 8. Interestingly, the machinist effect seems to increase in local wealth inequality. We suspect that this this effect is attributable to the presence of wealthy factory owners who utilized mechanized production methods in their establishments, requiring the intensive involvement of machinists. Alternatively, intergenerational mobility might have simply been lower in counties with higher concentration of wealth (Chetty and Hendren, 2018b; Chetty et al., 2014), which could make the catch-up of non-machinists' sons more difficult. Nonetheless, the interaction with public high schooling becomes even more negative in this column.[27] We conclude that the provision of public high schooling could dampen the difference between the offspring of machinists and non-machinists, in line with Becker and Tomes (1986). This result also suggests that machinist families did not have particularly strong preferences for education, since otherwise they could have sent their sons to college using money saved from substituting private with free public high school education or, simply, let their sons stay in public high school longer.

Third, being able to decipher blueprints, having some elementary knowledge of algebra or chemistry, and mechanical drawing skills were all valuable on the labor market in the late nineteenth century (Goldin and Katz, 2000, 2008). The sons of machinists could easily learn many of these skills from their fathers, thereby gaining some advantage outside formal schooling - which we call the information channel. However, schools increasingly started to incorporate scientific subjects into their curriculum which may have decreased the benefits of machinists' sons. To test this hypothesis, we use the Reports of the Commissioner of Education. These volumes contain relevant information - if the given school taught mechanical drawing or had a chemical laboratory - on two types of private high schools: institutions for secondary instruction and preparatory schools. We calculate the share of high school students whose school replied with a yes to any of the two questions. The underlying assumption is that these institutions put an emphasis on technical education in their curriculum. The interaction between technical education at school and a machinist father is estimated in Table 3.11. In line with our hypothesis, offering technical education decreased the relative gains accruing to machinists' sons, but only if private high schools were accessible to «rival» boys too (relatively low tuition fee; Column 2). Moreover, this effect is particularly strong in cities, where the benefits of technical skills could be reaped in manufacturing production, as opposed to rural areas and is also present for other potential outcomes (Columns 3-5).

[27] The highly significant interaction term in Column 3 of Table 3.9 also survives the inclusion of these control interactions.

Table 3.11: Information channel - sons (G2; 1900)

	Full sample	Low tuition fee (below city median)	Low tuition fee & population > 5.000 (city in 1870)		
	(1)	(2)	(3)	(4)	(5)
	Education rank	Education rank	Education rank	Max. primary education (% in occupation)	Sobek log-score
Machinist (G1)	4.502***	5.691***	3.846**	-2.277**	0.059*
	(0.787)	(0.894)	(1.449)	(0.979)	(0.032)
Technical education (% of HS students) x Machinist (G1)	-0.643	-1.852**	-3.489**	1.616*	-0.049**
	(0.652)	(0.781)	(1.458)	(0.945)	(0.019)
Manufacturing emp. (%) x Machinist (G1)	-0.065	-0.621	0.862	-0.479	0.013
	(0.539)	(0.888)	(1.499)	(0.868)	(0.028)
Baseline controls	Yes	Yes	Yes	Yes	Yes
County-fixed effects (1870)	Yes	Yes	Yes	Yes	Yes
Sample size	34475	20746	8327	8327	8070
Number of clusters	45	45	45	45	45

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1900) level. None of the specifications is weighted. The sample includes all sons who were not older than ten years in 1870 and whose father held an occupation between the 24.7th and 84.7th education rank percentiles in 1870. The sample is additionally restricted to sons who lived in 1870 i) in a county with private tuition fee below the median of cities (places with more than 5.000 inhabitants in 1870 - Col. 2-5) and ii) in places with more than 5.000 inhabitants in 1870 (Col. 5). The outcome variable is the education rank of occupation (Col. 1-3), the share of workers who had at most primary education in the son's occupation (Col. 4) and the Sobek log-score (Col. 5). The share of technical education (% of private high school students - institutions for secondary instruction or preparatory schools - whose school had a chemical laboratory or taught mechanical drawing) and of manufacturing employment (as % of county population in 1870) are winsorized at the 99th percentile and standardized. Baseline controls are described in Appendix C.1.1. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A decomposition of gains in earnings A simple goal is set in this subsection: understanding and quantifying to what extent the earnings effects of machinists' sons are driven by rural-urban differences and education.

First, we split the nominal earnings effect between sons who resided in villages (settlements with less than 5.000 inhabitants) and in cities in 1870. The first column of Table 3.12 shows that rural machinists' sons had significantly larger earnings gains: they had on average 4.7 log-points higher nominal earnings scores relative to city-dweller machinists' sons. A possible explanation is that the sons of rural machinists, being more educated than their peers, migrated to urban places more intensively or they migrated with their father to these areas and, thereby, had access to better paying urban occupations (see Tables 3.3 and C2). Machinists' sons in cities, on the other hand, could have experienced a relative urban premium only if initially urban non-machinists' sons would have left cities for rural areas.[28]

In line with the previous interpretation, Column 2 shows that the entire higher probability of urban place of living effect can be attributed to sons of initially rural machinists. We can use the differential probability between the offspring of rural and city-dweller machinists to calculate the differential earnings effect which can be explained by rural-urban earnings differences. Ward (forthcoming) estimates that rural-to-urban migration led to a 30 log-point increase in the log-earnings score in the early-twentieth-century United States. Assuming that this figure accurately describes the average gains of machinists' sons derived from rural-to-urban migration, we can conclude that the majority of the 4.7 log-points difference can be explained by the differential relative probability of urban status

[28]Additionally, the magnitude of earnings losses from urban-to-rural migration was significantly smaller than gains from rural-to-urban migration (Ward, forthcoming).

$(2.6 = 0.085 \cdot 30)$.

In Column 3, the sample is restricted to villagers' sons who lived in counties with medium level tuition fee (see Table 3.9). In line with anecdotal evidence in Goldin and Katz (2000) and Goldin and Katz (2008), we find that the (secondary) education of rural sons was indeed the pathway to urban life. At mean private high school provision, the son of a villager machinist was 12% more likely to live in an urban place three decades later than a comparable non-machinist's son. However, this effect increases by 70% when private high school provision increases by one standard deviation. We believe that this result lends support to the interpretation that machinists' sons ended up in urban places at least partly because they were more educated.

Second, we want to understand to what extent the rest of the machinist effect $(2.5 = 7.2 - 4.7)$ can be explained by returns to education. In unreported results, we establish that less than the half of the 0.21-year-longer schooling (see Column 1 in Table C5) stemmed from longer primary schooling, while the majority was the result of secondary schooling. Taking the returns to schooling estimates of Goldin and Katz (2000), we calculate that two-thirds $(1.7 = return\ to\ high\ school + return\ to\ primary\ school = 10.3\% \cdot 0.13 + 4.8\% \cdot 0.08)$ of the remaining machinist effect was the result of more years of schooling.[29] Considering that Goldin and Katz (2000) argue that their returns estimated in Iowa (1915) might be a lower bound on returns to education and that our estimated 0.21-year-longer schooling might be a lower bound too (owing to endogenous attrition), we can attribute practically the entire remaining earnings effect to returns to schooling.[30]

3.5.3 Fathers in other demanded occupations

The main results section is closed by looking at the sons of fathers in other occupations which were already present around 1870 and also received a boost from technological innovations during the Second Industrial Revolution (see Mokyr, 1999b).

The first such occupational group contains fathers who were chemists, engineers (mainly civil or mechanical), or telegraph operators - all white-collar jobs. Column 2 in Table 3.13 shows that their sons might have experienced even larger benefits,

[29]We cannot analyze a heterogeneous years of schooling effect by initial urban status owing to the small number of sons in the 1940 sample.

[30]The returns to education of Goldin and Katz (2000) combine within and across occupations gains, whereas earnings scores-based estimates can exclusively capture the latter. The estimates of Feigenbaum and Tan (2020) - those based on income scores measured before the Great Compression (Goldin and Margo, 1992) - indicate that 60-70% of the effect of a year of education on individual wages is captured in the effect on occupational earnings scores (4.4% vs 2.6-3.1%; see Tables 7 and A.9 of Feigenbaum and Tan, 2020).

Table 3.12: The urban-rural gap in the earnings effect (G2; 1900)

	State-level nominal log-score (1892)	Urban place of living (Yes=1)	
	(1)	(2)	(3)
	Full sample	Full sample	Villagers & medium tuition fee
Machinist (G1)	0.072***	0.099***	0.124***
	(0.019)	(0.023)	(0.032)
City (1870) x Machinist (G1)	-0.047*	-0.085***	
	(0.025)	(0.030)	
Private high school (%) x Machinist (G1)			0.088***
			(0.026)
Manufacturing emp. (%) x Machinist (G1)			-0.037
			(0.024)
Baseline controls	Yes	Yes	Yes
County-fixed effects (1870)	Yes	Yes	Yes
Sample size	45605	45605	6542
Number of clusters	45	45	45

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1900) level. None of the specifications is weighted. The sample includes all sons who were not older than ten years in 1870 and whose father held an occupation between the 24.7th and 84.7th education rank percentiles in 1870. The sample is additionally restricted to sons who lived in a place with less than 5.000 inhabitants and in a county with private tuition fee between the 25th and 75th percentiles in 1870 (see Table 3.9). The outcome variable is state-level nominal log-score (Col. 1) and an indicator for an urban place of living in 1900 (Col. 2-3). The specifications in Columns 1-2 also include a city indicator which equals to one if a son lived in a place with more than 5.000 inhabitants in 1870. Baseline controls are described in Appendix C.1.1. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

as measured by the education rank of occupation, than the sons of machinists. The conclusion is similar for the point estimate of the log-earnings score (Column 6), though this coefficient is imprecisely estimated, potentially owing to the small sample size.

Subsequent columns investigate the effect on the sons of two other, relatively lower-skilled groups of workers: employees of the railways (for instance, locomotive engineers or firemen) and operatives of the metal industry (smeltermen, heaters, etc.). Interestingly, we do not find evidence on any significant effect on their sons using our baseline matching estimation. While the explanation of the missing effect is beyond the scope of this work, we suspect that the labor market competition stemming from masses of low-skilled, European immigrants might have affected these lower-skilled workers more severely. Thus, the labor supply could more easily match the rising demand in these occupations.

3.6 Robustness checks

We discuss the robustness of our main findings below, implementing modifications in our baseline matching or regression estimations. In most cases, we concentrate on the effect on the two crucial outcomes of sons for the sake of brevity

Table 3.13: Sons of fathers in other occupations (G2; 1900)

	Education rank				Sobek log-score			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Machinist (G1)	2.403***				0.041***			
	(0.745)				(0.013)			
White-collar occupation boosted by the Second Ind. Rev. (G1)		4.959***				0.055		
		(1.757)				(0.038)		
Employee of railways (G1)			-0.300				0.030	
			(1.531)				(0.021)	
Metal industry operative (G1)				-0.517				-0.051
				(1.539)				(0.040)
Unbalanced controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	8745	1643	3317	2893	8424	1448	3020	1453
Number of clusters	45	45	43	43	45	42	44	41

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1900) level. All specifications are weighted by weights obtained from propensity score matching described in the main text. The final sample includes 1842, 200, 763, 665, 2779, 175, 746, 628 (Col. 1-8, respectively) matched sons of machinists. To mitigate the imprecision caused by the small number of treated observations, ten controls are chosen for the sons of white-collar workers instead of the usual five. The outcome variable is the education rank of occupation (Col. 1-4) and the Sobek log-score (Col. 5-8). Unbalanced controls included in the regressions are characteristics whose mean between machinist and control fathers is still significantly different at 5% after matching. The Sobek score merged to the occupation of fathers (1870) is also included in the propensity score matching in Columns 5-8. White-collar occupations boosted by the Second Industrial Revolution are: chemists, engineers (IPUMS's harmonized *OCC1950* code between 41 and 49), and telegraph and telephone operators. Railway employees are: brakemen, locomotive engineers, locomotive firemen and switchmen. Metal industry operatives are: filers, furnacemen, heaters, grinders, polishers and smeltermen. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(the education rank and urban place of living indicator in 1900). However, we deviate from these outcomes in a few specifications owing to sample size considerations and, instead, look at some outcomes of fathers.

3.6.1 Robustness checks using matching

Occupational employment pre-trends Facing occupational choice in their teenager years, fathers could elicit information about the future of certain occupations from their employment growth. For instance, the employment share of sailors was on a constant decline after the spread of steamships, indicating a gloomy future for prospective sailors. If machinists followed a relatively faster employment growth path compared to baseline control occupations, the identified positive effects could be the result of better foresight (and correlated talent) of machinist fathers, or simply the result of pre-trends leading to better occupation-level outcomes even in the absence of the Second Industrial Revolution. To assess this potential bias, we constructed the changes in the employment share of occupations at the census division level for the 1850s and 1860s (see Appendix C.1.4). These two measures are also included in the propensity score matching implemented in Columns 1 and 2 of Table C8. In comparison with Columns 2 and 5 of Table 3.5, both coefficients (insignificantly) increase in magnitude. Consequently, differential employment growth trends in the decades when fathers chose their occupation cannot explain our findings.

Manufacturing control occupations One might be concerned that workers of the manufacturing sector might have been more open-minded to modernity than people employed in more traditional sectors and predisposed to benefit from the overarching industrial and urban transformation in the late-nineteenth-century US. If this was the case, our matching estimation would be upward biased as the baseline control group contains many workers outside of manufacturing as well (e.g., carpenters or teamsters). Therefore, the sample is restricted to fathers who were employed in durable or non-durable manufacturing in Columns 3 and 4 in Table C8. However, this restriction causes no meaningful change in the coefficients of interest.

Maternal observables As more than 95% of mothers were not active on the labor market in our sample in 1870, we cannot use occupation-based measures of their socio-economic status. However, next to maternal age, we constructed an indicator variable if the mother was native born and if she was literate. Our baseline matching strategy in Section 3.5 balances our sample on maternal age and nativity even without including them as controls, but it is significantly more likely that a machinist's son had a more literate mother (2.6 percentage points difference - which amounts to a 9.6% standardized difference). We assess if more educated mothers drive our results in Columns 5 and 6 (Table C8), where we match on the three maternal observables as well. Once again, our findings are not affected by this change in the baseline specification.

Influential control occupations A particular concern could be an influential role played by the largest control occupations (Table C1). The interpretation of our findings would be profoundly different if the results were driven by a certain small group of control occupations. Therefore, we exclude fathers employed in the three largest control occupations - exclusively these three have a larger than five percent share among matched controls - from the pool of potential control individuals in last two columns of Table C8. The omission of these occupations, which provide approximately one-quarter of the control individuals in the baseline matching, does not influence the results in any significant way.[31]

The role of next-door neighbors The important effect of the neighborhood where kids grow up is well-established both in current and historical US context

[31] While the omission of the largest control occupations does not matter for our results, if control occupations experienced an employment decline or rise in 1870-1900 does matter. Restricting control occupations only to those which experienced an increasing (decreasing) employment share in these decades would result in different coefficients: 1.6 (4.3) for the educational rank and 0.034 (0.084) for the urban status indicator. In our baseline matching strategy, the average employment share change of the matched control group is approximately zero (unreported results).

(see e.g., Abramitzky et al., 2021; Chetty and Hendren, 2018a,b; Chetty et al., 2014; Durlauf, 2004; Galster, 2012; Ward, 2020). In our (subsequent) regression analysis, at most county-fixed effects can be included to capture the effect of growing up in the same neighborhood. However, within-county residential segregation along ethnic (Eriksson and Ward, 2019) or other socio-economic lines calls into question whether neighborhoods should be defined at the county-level.

To demonstrate that our results are not driven by machinists residing in more prosperous neighborhoods, we exploit the fact that next-door neighbors can be identified in the full count census. We construct the average value of personal property, real estate value, occupational education rank, literacy and foreign-born status of the closest household heads in 1870 (see Appendix C.1.7 for details). Reassuringly, our baseline matching strategy balances on these initially significantly different characteristics even without their inclusion (e.g., in the estimation in Table 3.5 - not reported). Thus, machinists tend to have very similar neighbors compared to matched control observations. Therefore, we believe that omitted differences in neighborhood quality cannot drive the findings.

The role of grandparental (G0) characteristics Grandfathers (G0) could influence our results and their interpretation in many ways. For instance, grandfathers with better foresight could nudge fathers to choose an occupation that was expected to be prosperous or to leave agriculture. Additionally, if machinists had significantly richer or more educated parents, this could introduce a more mundane form of omitted variable bias into the empirical analysis. All these reasons make the linking of fathers (G1) to their fathers (grandfathers; G0) important. In the resulting sample, we can assess the difference in coefficients with and without controlling for a large number of grandparental observables measured in 1860. Before doing so, we acknowledge that our sample might be selected since the parents of most foreign-born individuals did not live in the United States and some grandfathers might have died before 1860. However, the resemblance of coefficients estimated in Tables 3.3 and C9 suggests that the degree of this selection is not severe.[32]

First, we investigate how well the baseline matching strategy performs *without* explicitly balancing the sample on grandparental observables. The fact that the age, wealth (both real estate and personal property), urban status, population of place of living, and steel and iron industry dummy of grandfathers are significantly different before, but not significantly different after matching lends credibility to our estimation strategy. Furthermore, even when the difference cannot be eliminated in the case of certain remaining variables, it shrinks substantially. For

[32]The sole qualitatively different result is long-distance migration. Unlike the baseline analysis, where it is insignificantly positive, the coefficient becomes significantly positive at 5% in the new sample.

instance, non-machinist grandfathers are fifteen percentage points more likely to have an agricultural occupation initially. This gap is reduced to five percentage points with a p-value of 1%. Nonetheless, there are several variables which are still highly significantly different, the most prominent one being the indicator variable of a machinist grandfather.

Second, Table C9 reports the results with and without controlling for grandparental characteristics (see the notes below the table for the full list). It can be observed that the inclusion of these G0 background variables in the matching procedure does not change the results. Consequently, we can conclude that the main findings are not driven by grandparental observables.

3.6.2 Robustness checks using regressions

As a validation step before presenting the full set of robustness checks with fixed effects regressions, the baseline results for sons are estimated using these regressions instead of matching. The comparison of Tables 3.5 and 3.6 to Tables C6 and C7 reveals that the two estimation methods produce very similar coefficients which are not significantly different from each other.

Spatial sorting before 1870 Even though we can include county-fixed effects in our regressions, individuals who resided in a certain county in 1870 might have still been different in their migration history. Ideally, people who were born in a given county should not be compared to people who migrated there. Since the Second Industrial Revolution does not have a well-defined starting date, it could be the case that, when only county-fixed (1870) effects are used, in-migrated machinists with a good instinct to spot places with a growth potential are compared to locals who happened to be born there.[33]

Therefore, more detailed fixed effects are specified to tackle the possible spatial sorting prior to 1870. To do so, we generate fixed effects combining state of birth (country of birth for the foreign-born), county of living in 1870, an urban status indicator in 1870, and an indicator variable for above median age of the father. For instance, if a 28 year-old machinist was born in South Carolina, but then moved to the rural part of Erie county (NY), we are going to compare him to individuals with exactly the same migration history and below median age. Consequently, we will cease to compare individuals to all other locals in 1870. The underlying assumption is that individuals sharing the same migration history had very similar information and keenness to migrate. While the coefficients in Table C10 (Columns

[33]Klein and Crafts (2020b) argue that in the early-twentieth-century United States «*technological progress accelerated at this time but its progress was quite erratic and the development of new technologies and industrial locations was unpredictable.*»

1-2) somewhat decrease compared to Table C6, a large part of this insignificant difference is attributable to a slightly different, reduced sample.[34] This sample size reduction is the result of our narrowly defined fixed effects as we lose observations in less densely populated, rural areas or with a peculiar migration history. Finally, we can conclude that spatial sorting preceding the 1870s does not drive our results.

Additional state-occupation level pre-trends Our baseline matching strategy contains merely two occupation-state level characteristics (probability of migration and occupation change in the 1860s) because the matching algorithm would not convergence if many more were added. This limitation is simply the result of the occupation-based «treatment». However, many other similar variables can be included in fixed effects regressions. To this end, we calculated the two aforementioned variables for the 1850s, and added the average change in the urban status indicator and the probability of switching to an agricultural occupation for every occupation in the 1850s and 1860s (see Appendix C.1.5). The absence of any significant change after the inclusion of these control variables in Table C10 shows that the results are not outcomes of spatially-varying, occupation-level pre-trends.

Weighting for a representative sample The implementation of propensity score matching does not allow us to use any kind of weights. However, it is a well-known issue in the literature using the full count census that linking across different census waves might engender a non-representative sample. Therefore, we calculated the widely used inverse proportional weights to make the sample representative of the US population around 1870 (see Appendix C.2 for the details), then applied them in Columns 5-6 of Table C10. One can clearly see that our regression estimation without weighting produces coefficients very close to these new estimates. Therefore, we believe that our results accurately reflect the US population at the onset of the Second the Industrial Revolution.

Restricting the set of control occupations The baseline regression estimation includes all fathers whose occupation is above the 25$^{\text{th}}$ but below the 85$^{\text{th}}$ educational rank percentile. In the last robustness exercise reported in Table C10, we further restrict the sample of fathers to the 45$^{\text{th}}$-65$^{\text{th}}$ educational rank percentiles. No significant change ensues aside from a slight drop in the coefficients.

Old and young fathers/sons Before occupations start to grow rapidly or are about to decline, there is much uncertainty about their future. More forward-looking and able individuals might have anticipated the eventual rise of machinists

[34]Results with the new sample but without the new fixed effects are available upon request.

and took up this occupation early on. This type of sorting would imply that more positive effects should be observed for the sons of older machinists. Table C12 presents a comparison of the effect on sons depending on the age of the father. The age of sons is restricted between 0 and 5 in 1870 because otherwise older fathers have substantially older kids who, in turn, grew up in different years. Reassuringly, we do not find any significant difference between the sons of older and younger machinists when the sample is split by the age of the median machinist father.

Dynamic complementarity in the production of human capital is a well-established finding in the literature of education economics (see Caucutt and Lochner, 2020; Heckman and Cunha, 2007; Lee and Seshadri, 2019). This implies that those sons of machinists who were relatively old in 1870 should have experienced a relatively smaller increase in their level of education compared to the younger ones because they lacked complementary education investments during their early childhood. We investigate this question in the last column of Table C12. Confirming the theoretical prediction, machinists' sons who were older than ten years around the onset of the Second Industrial Revolution did not enjoy any gains in education (proxied by the education rank) in comparison with sons of similar, non-machinist workers.

3.6.3 Grandfather-fixed effects

Our arguably most important robustness checks are regressions in which grandfather-fixed effects are included. In other words, we compare machinists to their non-machinist brother(s). In this way, we can eliminate concerns related to machinists growing up in more advantaged families (unobservables not captured by the job, place of living or wealth of the grandfather) or inheriting a particular genetics, which helps them succeed in life (see Mogstad and Torsvik (2021) for a recent survey on this topic). To eliminate within-family differences in talent across siblings, we still control for many of their personal characteristics in 1870: county of living, education rank, literacy or wealth. Our regression specification thus takes the following form:

$$y_{f,c,g,1900} = \beta \cdot Machinist_{f,1870} + \gamma \cdot x'_{f,1870} + \delta_{c,1870} + \kappa_g + \epsilon_{f,c,g,1900} \tag{3.2}$$

where the fixed effect for grandfather g of father f appears as a new control variable (κ_g). Unfortunately, we can only apply this estimation strategy for fathers' outcomes due to sample size limitations. Moreover, the baseline sample must be extended in two ways even for fathers. First, we include all fathers who were between 16 and 50 years old (originally 20-40). Second, loosely defined occupations are not omitted anymore (e.g., 'Clerical and kindred workers (n.e.c.)').

Nevertheless, the baseline regression sample restriction is still implemented and we only include fathers whose occupation had an occupational education rank between 25$^{\text{th}}$ and 85$^{\text{th}}$ percentiles, thereby excluding farm laborers, fishermen but even high-skilled individuals such as bookkeepers or physicians.

The results of this estimation are shown in Table 3.14. The comparison of Columns 1 and 2 shows that the inclusion of grandfather-fixed effects does not significantly change the coefficient of interest in spite of a forty percentage-point increase in the R^2. This suggests that machinist fathers were significantly less likely to change their occupation even compared to their non-machinist brothers. In Column 3, the age of fathers is restricted to the original 20-40 range. The point estimate is practically unchanged but less precisely estimated owing to the sample size reduction. Next, we include all brothers irrespective of their education rank in Column 4. This produces an even larger coefficient than the initially estimated one in Column 2. Other outcomes of fathers are presented in Columns 5-7. The same conclusion can be drawn quantitatively and qualitatively as before (see Tables 3.3 and C9): a substantial positive likelihood of living in an urban place and (if anything) a positive probability to migrate across states.

The within-family estimation can greatly reduce the role of certain confounding unobservables, but it cannot entirely eliminate differences stemming from the different ability of brothers. In our previous analysis, we already made two steps to reduce their role. First, analogously to Feigenbaum and Tan (2020), who restrict their sample to small years of education differences between twins, brothers holding occupations with the lowest and highest education ranks were excluded. The underlying assumption is that brothers with more similar education ranks are more likely to be similar in terms of unobservables as well. Second, the included personal characteristics (for instance, the two wealth measures, the education rank of occupation or literacy dummy) should already capture a certain degree of differences in ability. To further reduce the likelihood that the results are driven by unobserved ability, we borrow from the literature which estimates returns to schooling using twins (e.g., Ashenfelter and Rouse, 1998; Feigenbaum and Tan, 2020). They argue that some observable variables - marriage status,[35] spousal education, number of kids, etc. - are correlated with ability. In Table C11, we demonstrate that none of these variables are correlated with the machinist dummy. Perhaps even more importantly, specifications *without* grandfather-fixed effects show no significant association either.

[35]In the absence of a separate census question on marriage status in 1870, a father is imputed to be married if the age of the spouse is known.

Table 3.14: Within-family estimation - fathers (G1; 1870-1900)

	Occupational change (Yes=1)				Other outcomes		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	Baseline	20-40 y.o.	Unrestricted sample	Urban in 1900 (Yes=1)	Migration (within-state; Yes=1)	Migration (across states; Yes=1)
Machinist (G1)	-0.107***	-0.138**	-0.133*	-0.204***	0.116**	-0.027	0.066
	(0.037)	(0.062)	(0.079)	(0.041)	(0.050)	(0.043)	(0.049)
Grandfather (G0)-fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes
County-fixed effects (1870)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Personal controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	22799	22799	17793	69221	22799	22799	22799
R^2	0.24	0.64	0.65	0.67	0.72	0.62	0.65

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are multiway clustered at the grandfather-county (1900) level. None of the specifications is weighted. The sample includes all fathers who held an occupation between the 24.7th and 84.7th education rank percentiles in 1870 - except for Column 4 which includes all fathers irrespective of the education rank of their occupation. In every column, the age of included fathers is between 16 and 50 years (inclusive) - except for Column 3 where the age is restricted between 20 and 40 years (inclusive). The outcome variable is a binary variable which equals one if i) the father changed occupation between 1870 and 1900 (Col. 1-4), ii) the father lived in an urban place in 1900 (Col. 5); iii) the father migrated within-state across counties (Col. 6) or across states (Col. 7). Personal controls included in the regressions are (all measured in 1870): the education rank of occupation, urban status and literacy indicator, age (in years), value of real estate and personal property, number of inhabitants in the place of living and a farmer-farm manager-farm foreman indicator. The interactions of the urban indicator, size of place of living, two wealth measures, education rank and age are also included. The squared size of place of living, wealth measures and age are included as well. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.6.4 Correcting measurement error and magnitude comparison

It is well-known that the misreporting of binary independent variables produces a non-classical measurement error in regression estimations because the measurement error is mechanically negatively correlated with the correctly measured value (see e.g., Aigner, 1973; Bingley and Martinello, 2017; Dupraz and Ferrara, 2021). Consequently, the OLS estimate is a lower bound on the consistent coefficient normally. The relationship between the correct coefficient and inconsistent OLS estimate is the following:

$$plim\ \hat{\beta}_{OLS} = \beta \cdot (1 - p - q) \tag{3.3}$$

where β is the consistent coefficient, p is the share of false positives (among fathers classified as machinists, $p\%$ were incorrectly classified as one), and q is the share of false negatives (among fathers classified as non-machinists, $q\%$ were actually machinists).

In our case, q can be set equal to zero owing to the small share of machinists in the whole sample. A non-negligible p can be the result of two, distinct measurement errors. First, a machinist observation might be linked to a non-machinist one when we link across census waves. For conservative linking methods used in this paper, Bailey et al. (2020) estimate a false positive ratio of 10-15%. Second, even if we could perfectly link individuals to their own observations over time, the misreporting of occupations can cause measurement error. Ward (2019) shows that around one-third of respondents misreported their occupation in the full count census, relying on a census re-enumeration in Saint Louis in 1880.[36] Therefore, we believe that assuming $p \approx 40\%$ might capture the true extent of false positives.

Using the previously introduced formula, one can see that the OLS coefficient is assumed to be downward biased by a factor of 0.6 (=1-0.4). Under this assumption, the consistently estimated effects are around 66.67% larger than the earlier OLS estimates. This implies that a machinist's son had on average a four percentiles higher education rank (Col. 2 of Table 3.5), 0.35 years more of schooling (Col. 1 of Table C5), and a seven log-points higher nominal earnings score (Col. 2 of Table 3.6) than a son of a comparable but non-machinist father.

3.7 Conclusion

In this paper, we investigate to what extent and how individuals in occupations that are beneficially affected by structural transformations can transmit their gains

[36] If a reported machinist was more than 66.67% likely to actually hold the machinist occupation, the magnitude of the adjustment factor declines along with p.

in socio-economic status to their offspring. Combining full count census data with newly digitized data sources, we establish that machinists, whose occupation experienced a relative labor demand spike in the United States during the Second Industrial Revolution, experienced relatively higher income and job stability. Relying on propensity score matching and fixed effects regressions, we document that the (grand)sons of machinists were significantly better-off in terms of earnings-related outcomes than (grand)sons of observationally similar non-machinists. In addition, the main contribution of this work is pinning down the mechanism which underlies the documented intergenerational transmission. We find that the sons of rural machinists benefited from rural-to-urban migration and parental investment in their education, while the sons of urban machinists mostly gained from the latter channel. A wide range of robustness checks show that the results are unlikely to be driven by the (transmitted) unobserved ability of machinist fathers.

In conclusion, the main mechanisms behind intergenerational mobility seem to have changed little over more than a century: the opportunities offered by high-quality urban neighborhoods (see Chetty and Hendren, 2018a,b; Chetty et al., 2014; Durlauf, 2004; Galster, 2012; Laliberté, 2021) and by high educational attainment guarantee a higher socio-economic status in the age of telegraphs as well as of smartphones. We also show that expanding public schools could equally well reduce inequality stemming from financially constrained parents in the past as nowadays (Dobbie and Fryer, 2011; Duflo, 2001; Lucas and Mbiti, 2012; Neilson and Zimmerman, 2014; Wantchekon et al., 2015). Taken together, our results suggest that the effects of current transformations in the labor market, such as automation, might be passed on to later generations, but to a lesser extent due to today's considerably more expanded public education and unemployment benefit system (allowing for less occupational downgrading) - especially if people are allowed to move to places offering better economic prospects.

C Appendix - Chapter 3

C.1 Data appendix

C.1.1 Controls used in propensity score matching

The controls used in the baseline propensity score matching are the following. For every father in the census in 1870, we measure:

- Personal characteristics: age (in years), literacy (can read and write, yes=1), foreign-born dummy (yes=1), native-born dummy (yes=1), dummies for the UK (yes=1) and for Germany (yes=1) as country of birth;

- Occupational characteristics: education rank of occupation (percentile rank; Song et al., 2020),[37] occupation-state level migration and probability of occupation change between 1860 and 1870 (see Appendix C.1.5);

- Measures of individual wealth: value of personal property and real estates (separately);

- Characteristics of place of residence: urban status (yes=1), size category of place of living,[38] dummy for living in the state of birth (yes=1) and state-fixed effects.

Moreover, we use the pairwise interactions of the following six variables: real estate, personal property, age, urban dummy, population of place of living, education rank. We also include the square of the two wealth measures, the age and the population of place of living. Finally, we include several county characteristics downloaded from the NHGIS (Manson et al., 2021; for 1870): the share of manufacturing employment (% of total population), manufacturing output per capita, manufacturing output per manufacturing wage earners, the share of steam engine-provided engine power (% of steam engine- and water-driven engine-provided total). We refer to these controls jointly as *baseline controls*.

C.1.2 The construction of state-level earnings scores

The source of state-level earnings data is the *Fifteenth Annual Report of the Commissioner of Labor* which reports daily average wages for US states and other countries mainly for years in the second half of the 19th century. We sought to find a close match for every occupation i) which has a large role as control occupation for machinist fathers in 1870, or ii) which is a common occupation across fathers or sons in 1900.

We checked all relevant state-occupation pairs for 1870-72, 1879-1881 and 1890-1892, and digitized every entry in which at least ten individuals were used for average wage calculation. In case of multiple entries within any of the three-year time spans for a given state, we chose the average wage which was based on the largest number of wage reporting individuals. We exclusively included entries for males. The ending years of 1872, 1881 and 1892 were chosen because they preceded the Panics of 1873 and 1893, and the Depression of 1882-85. In a few cases, we deviated from our baseline data collection strategy to improve our sample. For miners, census-based, daily average wages were used from 1889 for

[37] We use the first available rank which is constructed for those born around 1880.

[38] We converted the original *SIZEPL* variable into actual population numbers using the midpoint of every interval. The first and last categories are defined using half the length of the second and penultimate intervals, respectively.

the 1892 income score because the number of reporting states and observations used for average wage calculation were undoubtedly superior to other publications between 1890-1892. For farm laborers, data were digitized from the *Ninety-ninth Bulletin of U.S. Department of Agriculture (Wages of Farm Labor).* Daily wages were digitized *without* board (accommodation) to reduce the gap in in-kind compensation between agricultural and manufacturing laborers (Alston and Hatton, 1991, Hatton and Williamson, 1991). For the 1892 score, the number of occupations available is increased in our sample by using the publication titled *The slums of Baltimore, Chicago, New York, and Philadelphia: prepared in compliance with a joint resolution of the Congress of the United States (1892).* We take mostly occupations in services[39] for Maryland, Illinois, New York and Pennsylvania.

Next, daily wages were converted to yearly earnings following Sobek (1996). We assumed 245 days of work for the majority of occupations, 225 days for building trades (bricklayers, cabinetmakers, carpenters, masons, painters, plasterers) and farm laborers, 270 days for clerical occupations (bookkeepers, clerks, telegraph operators).[40] For farm labor, we assumed 30 days of harvest wages and 195 (=225-30) days of non-harvest wages.[41] We multiplied the yearly earnings of farm labor by the ratio of farmer-to-farm labor score in Sobek (1996) to compute earnings scores for owner-occupier farmers.

The main limitation of our earnings score is that the earnings of high-skilled workers, for instance, lawyers or physicians, cannot be observed. To solve this problem, the following imputation procedure is set up. First, we took the earnings scores of the fifteen occupations (TOP15) for which we have the most state-year level observations.[42] Afterwards, we calculated the earnings scores of missing, predominantly white-collar occupations[43] by multiplying our earnings scores for

[39]These occupations are: barbers, bartenders, watchmen, policemen, detectives, agents (n.e.c), clerks, longshoremen, hucksters, salesmen (n.e.c). We included an observation if the average wage could be calculated using at least ten individuals.

[40]The slum report provides weekly wages. Following Sobek (1996), we assumed 45 weeks worked apart from clerical jobs, where 48 weeks are assumed.

[41]Unlike Sobek (1996), we did not assume 245 days of work for farm laborers because it gave rise to a tendency of *nominal* farm laborer wages surpassing laborer wages. This would be inconsistent with existing evidence (Alston and Hatton, 1991; Hatton and Williamson, 1991). Our ratio between farm laborer to laborer nominal earnings scores is really close to the estimates found in the literature which takes into account the pecuniary value of in-kind remuneration as well. Moreover, the similar length of (un)employment spells between farm workers and workers in building trades is also consistent with Engerman and Goldin (1991).

[42]These occupations are: blacksmiths, boilermakers, cabinetmakers, carpenters, compositors, engineers (locomotive), firemen (locomotive), laborers (n.e.c.), machinists, molders, painters, pattern makers, plumbers, stone cutters and teamsters.

[43]These occupations are: operatives (n.e.c.), managers, physicians, lawyers, meat cutters, clergymen, pharmacists, policemen, insurance agents, foremen (n.e.c.), teachers (n.e.c.), craftsmen (n.e.c.), fishermen, engineers (civil and mechanical separately), accountants, chemists, draftsmen,

the available TOP15 occupations with the ratio of the Sobek score of the missing occupation and each of the TOP15 occupations. Then, we took the unweighted average of the implied earnings scores which constitutes our earnings score for missing occupations. We calculated this average only if at least eight of the fifteen (more than half) occupations were available for the case of a given state-year pair in order to reduce measurement error. The main assumption underlying this imputation procedure is that the ratio of earnings scores found in Sobek (1996) around 1890 is the same across states (for our 1892 score), or the same across states and time (for our 1872 and 1880 scores). Reassuringly, Katz and Margo (2014b) find that the skilled artisans-to-clerks earnings ratio remained stable between the 1840s and 1880s. The debate if the earnings of higher-skilled workers differed across states more or less than the earnings of production workers or craftsmen has not been settled yet (see Goldin, 1998; Rosenbloom, 1990, 1996, 2002; Sundstrom and Rosenbloom, 1993). Therefore, applying the Sobek score ratio-implied premia for earlier decades might not introduce a large measurement error since most of our TOP15 benchmark occupations are classified as artisans/craftsmen and we mostly impute the wages of white-collar workers.[44]

Another empirical barrier is that some harmonized occupations have many potential matches in our earnings score data. For instance, we had to aggregate the earnings score of miners of coal, iron or zinc into a single score for miners. The affected occupations are brickmasons (bricklayers and masons), railroad conductors (freight, passenger or not specified), miners (coal, iron, lead and zinc), spinners and weavers (cotton or woolen goods). Our state-year level earnings score for these harmonized occupations is defined as the observation-weighted average earnings score (of «subcategories»).[45] As a last step, missing earnings scores were imputed with the unweighted average of states within a given census division whenever it was possible.

The estimates of `measuringworth.com` were used to convert all earnings

editors, funeral directors, musicians, ship officers, stenographers, real estate agents, janitors, waiters, gardeners and sailors. For 1872 and 1880, the list also includes barbers, bartenders, agents (n.e.c.) and hucksters.

[44]For 1872, we need an additional step because there are no data on the wages of miners, shoemakers and tailors who play an important part in the control group of machinists. To impute their wages, we follow our procedure described in the main text with one exception. Instead of using the ratio of Sobek scores, we calculate our observation-weighted, US-level earnings score in 1880 for the TOP15 occupations as well as for miners, shoemakers and tailors. We use the ratio of these earnings scores to implement the imputation procedure in order to diminish the potential effect of the Second Industrial Revolution on relative wages over time.

[45]Additionally, we included furnacemen in foundries or in the gas industry as furnacemen, and lumbermen can be lumber handlers, lumber pilers or wood choppers as well. Two of the different «subcategories» of furnacemen or lumbermen never coincided within a state-year cell. Thus, there was no need to calculate observation-weighted averages.

scores into 1890 dollars. The 1872/1880 earnings scores were multiplied by 0.75/0.89.

The conversion of nominal earnings scores to real scores requires state-level and urban-rural price differences. Nominal earnings (1872, 1880 and 1892) were deflated by the state-level price index of Haines (1989), and nominal wages (1940) by the cost of living measures reported in Stecker (1937). As Haines (1989) and Stecker (1937) do not contain information on all states, we use the price index of a neighboring state in the case of missing values (the actual pairs are available upon request).[46] We inflate earnings scores in places with less than 25.000 inhabitants by 1.192 (1872, 1880 and 1892 - Hatton and Williamson, 1991) and by 1.205 (1940 - Williamson and Lindert, 1980) to account for urban-rural price differences. In doing so, we follow the best practice in earlier literature (e.g., Collins and Wanamaker, 2014).

C.1.3 The construction of schooling supply measures

The source of our high school supply proxies are different *Reports of the Commissioner of Education*. We followed a distinct data collection strategy for private and public high schools.

Private high schools We refer to institutions for secondary instruction, preparatory schools, commercial and business colleges (excluding evening schooling), preparatory departments of colleges and universities, and schools of science as private high school.

First, all available data on private high schools were digitized from the 1880 Report. If a school was reported as not replying to the query of the Commissioner's office, we tried to find it in the 1882 Report. Different types of schools were expected to report different data, so the following pieces of information could be digitized:

- Institutions for secondary instruction: number of teachers and students (split by gender), tuition fee, dummy whether mechanical drawing is taught, dummy if they had a chemical laboratory;

- Preparatory schools: number of teachers and students,[47] tuition fee and dummy if they had a chemical laboratory;

[46] Stecker (1937) reports cost of living for more than one city in some states. We calculated the unweighted average of cost of living in cities within those states.

[47] Preparatory schools are not included in our high school student shares since we do not know the exact number of male students.

- Commercial and business colleges: number of teachers and students[48] (split by gender), and tuition fee;
- Preparatory departments of colleges and universities: number of teachers and students (split by gender);
- Schools of science: number of teachers and students (split by gender).

If tuition fees were not reported for the entire scholastic year (but for a term or month), a 40-week (10-month) long scholastic year was assumed which was the most common length. The children of residents sometimes did not have to pay the tuition fee. In such cases, the tuition fee is set equal to zero. In the next step, schools were matched to counties (1870) one-by-one using their reported location (post office).

Public high schools The data collection process for public high schools is more complex. While detailed statistics were reported for private high schools starting from the 1870s, no school-level information is available on public high schools until 1890. Moreover, the year of establishment is solely recorded in the Reports published in the mid-1900s.

To circumvent these data limitations, we adopted the following data collection strategy. First, we restricted our attention to schools in cities which had a population of 7.500 in 1880 since municipality/school name changes between 1890 and the mid-1900s would be an insurmountable barrier to data collection considering the number of public schools. Then, we turned to Reports of the mid-1900s for the list of public high schools which were established in these cities until 1880. Next, all available data were digitized on these high schools in the 1890/91 Report. If a school did not report despite being established pre-1880, we searched for it in the 1892/93 Report. For high schools which existed in 1890 but had no establishment year, we searched the web to gather information about their establishment year. As a result of this process, we obtained information on the number of teachers and students (split by gender) in public high schools around 1890. We believe that this value should be strongly positively correlated with its counterpart in 1880 since high school completion rates started their rapid increase only after the turn of the century (Goldin, 1998; Goldin and Katz, 2008). One might also argue that in the largest, fastest growing cities schools might have been split between 1880 and 1890 and, consequently, we underestimate the true extent of high school provision. Nonetheless, we show in the relevant analysis that our results our robust to the omission of these metropolises.

[48] We digitized the number of students in day education if it was available. Otherwise, the missing value was imputed with the number of all students including evening schooling.

Imputation of missing values Before the creation of the final measures of schooling supply, missing values for the six school types had to be imputed. We followed the same procedure for all of them (except for public high schools - see the last paragraph of this section). First, if the number of all students was missing, we used the unweighted average of the same type of schools within-state (if there were less than ten such schools, then within-census division). The number of male students was imputed using the unweighted share of males in the same type of schools within-state (if there were less than ten such schools, then within-census division) and multiplying it by the (imputed) number of all students. The number of teachers was imputed similarly - the unweighted average of the same type of schools within-state (if there were less than ten such schools, then within-census division). Finally, a missing tuition fee was imputed as the number of students-weighted tuition fee within-state (if there were less than ten schools of the underlying type, then within-census division).

The aggregation of school-level measures to the county level amounts to a simple summation of the number of students and teachers, and taking the weighted average (by number of students) in case of the tuition fee. The five different private school types were pooled together before summation. The share of private and public high school students was calculated as the number of male students divided by the number of males aged 14-20 in a given county in 1880. The teacher-pupil ratio is defined as the student-weighted ratio of teachers to all students (male and female) at each school. The share of students having technical education (at institutions for secondary instruction or preparatory schools) was constructed as follows. All students who were at a school which offered mechanical drawing or had a chemical laboratory were indicated as having technical education. The sum of these students is divided by the total number of students at the county level.

For public high schools, the strong dependence of school size on local population necessitated a different imputation strategy. First, we ran the following regression:

$$y_{c,s} = \beta \cdot Population_{c,1880} + \gamma \cdot Population^2_{c,1880} + f_{state(c)} + \epsilon_{c,s} \tag{3.4}$$

where $y_{c,s}$ is the number of students or teachers in city c and public high school s. City population and its squared form (population figures are from the Report of the Commissioner of Education in 1880), and the state-fixed effects produce an $R^2 \approx 0.5$. This model is used to impute the missing number of students and teachers if a public high school already existed before 1880. To split the number of students by gender, the average gender ratio is used within state - if at least ten public high schools have non-missing data -, otherwise the average of public high schools in the census division.

C.1.4 Occupational employment growth until 1870

The full count censuses of 1850, 1860 and 1870 are used to compute changes in employment shares in the period preceding the Second Industrial Revolution. In all three years, we dropped individuals who were not between 16 and 65 years old and gave a non-occupational response (*OCC1950* codes larger than 978). Then, we calculated the share of every harmonized occupation for each census division. Last, we created the differences between 1850-60 and 1860-70, and merged them to fathers in 1870 based on their occupation and census division.

C.1.5 Occupation-state level measures in the pre-period

We computed several occupation-state level measures based on the 1850, 1860 and 1870 full count censuses. To do so, the census was first restricted to individuals between 16 and 40 years old. We assigned to every state the occupational level i) probability of changing occupation; ii) probability of migration (changing county or state), iii) average change in the urban status dummy, iv) probability of having agricultural occupation at the end of the decade - based on individuals who at the beginning of the decade (1850s and 1860s) lived in the given state.

C.1.6 Imputing self-employed income

We followed the literature in imputing the income of self-employed individuals in the 1940 full count census (see e.g., Collins and Wanamaker, 2017; Ward, forthcoming). As a first step, a sample of male self-employed workers was created in the 1960 5% census. We calculated the ratio between the total income and wage income for these individuals. Finally, the wage of self-employed individuals with non-zero reported wage in 1940 was inflated by the median of the calculated ratio (1.89). The main assumption of this imputation is that the ratio remained constant between 1940 and 1960.

For self-employed individuals, who reported zero wage earned in 1940, we use the median total income obtained from the 1960 census after a conversion from 1960 to 1940 dollars and conditional on reporting more than 50 weeks worked. We calculated the median separately for the agricultural (*OCC1950:* 100, 123, 810, 820, 830, 840) and non-agricultural self-employed.

C.1.7 Characteristics of next-door neighbors

To calculate observable measures for next-door neighbors, we first took every household head from the 1870 census. This data set is sorted, so neighbors appear next to each other. To every household head we assigned its ten closest neighbors, i.e. the five household heads right before and after a given person.

Afterwards, the real estate and personal property values were winsorized at the 1st and 99th percentiles. Finally, the average of observable neighbor characteristics was computed and they were assigned to the 1870 full count census. We use the occupational education rank estimated for the 1880 (earliest) cohort by Song et al. (2020) for all neighbors. Literacy (foreign-born status) is measured as a dummy which is set equal to one if a given neighbor could read and write (was born outside the US).

C.2 Inverse proportional weights

To create inverse proportional weights for the sons' sample, sons were linked between 1870 and 1900 with the two conservative linking methods developed by Abramitzky et al. (2020). First, we merged the full count census of 1900 to the crosswalk, keeping matched as well as unmatched observations. Next, we also merged this data set with the 1870 full count census. If an observation could not be matched with any of the two conservative linking methods, we considered it unmatched and generated a variable which was set to zero for this case (one otherwise). Then, we used this binary variable as an outcome of a probit regression on age bins (following the code provided by Abramitzky et al., 2020), an urban place of living indicator, the population size category of place of living (*SIZEPL*) and census division-fixed effects - all measured in 1900. Finally, the inverse proportional weight for every single matched observation was calculated based on the following formula: $(1-\hat{p})/\hat{p}$, where $\hat{p}$ is the predicted probability of a successful match. We set the weight equal to zero for observations which were unmatched.

C.3 Additional empirical results

Table C1: Top control occupations

Top control occupations (*OCC1950*)	**% of all control observations**
Carpenters	9,92%
Truck and tractor drivers	7,68%
Shoemakers	6,16%
Painters (construction)	4,17%
Blacksmiths	4,15%
Masons	3,34%
Hucksters and peddlers	2,94%
Stationary engineers	2,91%
Tailors	2,58%
Molders (metal)	2,36%
Bookkeepers	2,24%
Compositors and typesetters	2,10%
Meat cutters	1,99%
Stone cutters	1,86%
Clergymen	1,64%

Note: the results presented in this table pertain to the propensity score matching in Table 3.5.

Table C2: Migration destination decomposition - fathers (G1; 1900)

	Migration (within and across states) [(1)=(2)+(3)]			Urban destination [(2)=(4)+(5)]		Rural destination [(3)=(6)+(7)]	
	(1) Any destination	(2) Urban destination	(3) Rural destination	(4) Urban in 1870	(5) Rural in 1870	(6) Urban in 1870	(7) Rural in 1870
Machinist (G1)	0.017	0.037***	-0.020***	0.016	0.021***	-0.013*	-0.007**
	(0.011)	(0.009)	(0.007)	(0.011)	(0.004)	(0.006)	(0.003)
Mean of outcome	0.58	0.28	0.30	0.15	0.13	0.13	0.17
Standard deviation of outcome	0.49	0.45	0.46	0.36	0.33	0.33	0.38
Unbalanced controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	18811	18811	18811	18811	18811	18811	18811
Number of clusters	50	50	50	50	50	50	50

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1900) level. All specifications are weighted by weights gained from propensity score matching described in the main text. The summary statistics reported are unweighted and pertain to the full estimation sample before matching. The outcome variable is a binary variable which equals one if the father migrated between 1870 and 1900 (across or within states; Col. 1), if he migrated and was found in an urban (Col. 2) or rural (Col. 3) place of living in 1900, if he migrated to an urban destination by 1900 and lived in an urban (Col. 4) or rural (Col. 5) place of living in 1870, if he migrated to a rural destination by 1900 and lived in an urban (Col. 5) or rural (Col. 6) place of living in 1870. Unbalanced controls included in the regressions are characteristics whose mean between machinist and control fathers is still significantly different at 5% after matching. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C3: Measures of economic status - fathers (G1; 1880)

	(1) State-level nominal log-score (1880)	(2) State-level real log-score (1880)
Machinist (G1)	0.065**	0.057**
	(0.025)	(0.023)
Mean of outcome	6.07	6.20
Standard deviation of outcome	0.42	0.42
Unbalanced controls	Yes	Yes
Sample size	19120	19120
Number of clusters	47	47

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1880) level. All specifications are weighted by weights gained from propensity score matching described in the main text. The summary statistics reported are unweighted and pertain to the full estimation sample before matching. The final sample includes 4428 matched machinist fathers. The outcome variable is the state-level nominal and real log-score (Col. 1 and 2). Unbalanced controls included in the regressions are characteristics whose mean between machinist and control fathers is still significantly different at 5% after matching. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C4: Measures of occupational income and mobility - occupation switcher fathers (G1; 1870-1900)

	Earnings scores (state-occupation in 1900)					Occupational mobility measures (1900)	
	(1) Sobek log-score	(2) State-level nominal log-score (1892)	(3) State-level real log-score (1892)	(4) Higher-paying occupation (+$150 or more; Yes=1)	(5) Lower-paying occupation (-$150 or less; Yes=1)	(6) Manager/official/ proprietor (Yes=1)	(7) Siegel's prestige log-score
Machinist (G1)	0.036**	0.032**	0.026**	0.004	-0.060***	-0.015	0.028**
	(0.014)	(0.012)	(0.013)	(0.018)	(0.016)	(0.010)	(0.012)
Mean of outcome	6.07	6.17	6.30	0.22	0.34	0.13	3.56
Standard deviation of outcome	0.60	0.45	0.43	0.42	0.47	0.33	0.36
Unbalanced controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	7337	7337	7337	7337	7337	7337	7337
Number of clusters	47	47	47	47	47	47	47

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1900) level. All specifications are weighted by weights obtained from propensity score matching described in the main text. The summary statistics reported are unweighted and pertain to the full estimation sample before matching. The final sample includes 1578 matched machinist fathers and non-machinist fathers who did not hold the same occupation in 1870 and 1900. The outcome variable is the Sobek, state-level nominal and real log-score (Col. 1-3), an indicator variable if the state-level real log-score was at least $150 higher (Col. 6) or lower (Col. 5) in 1900 than in 1870, an indicator variable if the father held a managerial/proprietor occupation (Col. 6; *OCC1950* code=290), and Siegel's prestige log-score (Col. 7). Unbalanced controls included in the regressions are characteristics whose mean between machinist and control fathers is still significantly different at 5% after matching. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C5: Measures of education and wealth - sons (G2; 1940)

	(1) Highest grade completed	(2) Some primary education (years < 9; Yes=1)	(3) Some secondary education (9 <= years <=12; Yes=1)	(4) Some university education (12 < years; Yes=1)	(5) Owned a house (Yes=1)
Machinist (G1)	0.209**	-0.032**	0.036***	-0.004	0.018
	(0.091)	(0.012)	(0.010)	(0.009)	(0.019)
Mean of outcome	7.91	0.75	0.17	0.08	0.68
Standard deviation of outcome	3.42	0.44	0.38	0.28	0.47
Unbalanced controls	Yes	Yes	Yes	Yes	Yes
Sample size	7543	7543	7543	7543	7543
Number of clusters	49	49	49	49	49

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1940) level. All specifications are weighted by weights gained from propensity score matching described in the main text. The summary statistics reported are unweighted and pertain to the full estimation sample before matching. The final sample includes 919 matched machinist sons. We use ten matched control observations instead of five owing to the small number of machinist sons. The outcome variable is the highest grade of schooling completed (Col. 1 - winsorized at the 99th percentile), a binary variable which equals one if i) the years of schooling is below nine years (Col. 2), ii) the years of schooling is between nine and twelve years (Col. 3), or iii) the years of schooling is more than twelve years (Col. 4), and an indicator variable for house ownership (Col. 5). Unbalanced controls included in the regressions are characteristics whose mean between machinist and control fathers is still significantly different at 5% after matching. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C6: Main outcomes - sons (G2; 1900)

	Occupational characteristics		Migration (Yes=1)		Place of living (1900)			Personal characteristics		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Agricultural occ. (Yes=1)	Education rank	Within-state	Across states	Urban (Yes=1)	Higher population than in 1870 (Yes=1)	Manuf. emp. per capita (county)	# of children	Married (Yes=1)	Owning house (Yes=1)
Machinist (G1)	-0.025***	3.368***	-0.014	0.048***	0.044***	0.035***	0.003	-0.051	0.002	-0.001
	(0.007)	(0.648)	(0.010)	(0.011)	(0.009)	(0.012)	(0.002)	(0.040)	(0.009)	(0.011)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County-fixed effects (1870)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	63857	63857	63857	63857	63857	63857	63857	63857	63857	63857
Number of clusters	45	45	45	45	45	45	45	45	45	45

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1900) level. None of the specifications is weighted. The sample includes all sons whose father held an occupation between the 24.7th and 84.7th education rank percentiles in 1870. The outcome variable is an agricultural occupation indicator (Col. 1 - farmer, farm manager/foreman/laborer), the education rank of occupation (Col. 2), a binary variable which equals one if the son migrated within-state across counties (Col. 3) or across states (Col. 4) between 1870 and 1900, a binary variable which equals one if the son lived in an urban place in 1900 (Col. 5) or his place of residence fell into a larger *SIZEPL* category in 1900 than in 1870 (Col. 6), manufacturing employment per capita (as % of total county population), the number of children in the household (Col. 8), a marriage status (Col. 9) and house ownership (Col. 10) indicator. Baseline controls are described in Appendix C.1.1. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C7: Measures of economic status - sons (G2; 1900)

	(1) Sobek log-score	(2) State-level nominal log-score	(3) State-level real log-score	(4) State-level real score (level)	(5) Preston-Haines log-score
Machinist (G1)	0.060***	0.040***	0.030***	15.170**	0.075***
	(0.011)	(0.010)	(0.010)	(6.406)	(0.009)
Baseline controls	Yes	Yes	Yes	Yes	Yes
County-fixed effects (1870)	Yes	Yes	Yes	Yes	Yes
Sample size	49268	44553	44553	44553	40331
Number of clusters	45	45	45	45	45

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1900) level. None of the specifications is weighted. The sample includes all sons whose father held an occupation between the 24.7th and 84.7th education rank percentiles in 1870. The outcome variable is the Sobek, state-level nominal and real log-score (Col. 1-3), the state-level real score in levels (Col. 4) and the Preston-Haines log-score (Col. 5). Baseline controls are described in Appendix C.1.1. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C8: Robustness checks - sons (G2; 1900)

	Occupational pre-trends		Manufacturing control occs		Maternal characteristics		Top 3 control occs excluded	
	(1) Education rank	(2) Urban (Yes=1)	(3) Education rank	(4) Urban (Yes=1)	(5) Education rank	(6) Urban (Yes=1)	(7) Education rank	(8) Urban (Yes=1)
Machinist (G1)	4.133***	0.074***	2.913**	0.050***	3.153***	0.050***	2.107***	0.052***
	(1.010)	(0.015)	(1.210)	(0.015)	(0.830)	(0.013)	(0.641)	(0.012)
Unbalanced controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	7015	7015	3111	3111	8904	8904	8570	8570
Number of clusters	45	45	45	45	45	45	45	45

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1900) level. All specifications are weighted by weights gained from propensity score matching described in the main text. The sample includes 1842 (Col. 1-2, 7-8), 1794 (Col. 3-4) and 1823 (Col. 5-6) matched machinist sons. The outcome variable is the education rank of occupation (every odd column) or an urban place of living indicator (every even column). Unbalanced controls included in the regressions are characteristics whose mean between machinist and control fathers is still significantly different at 5% after matching. In Columns 1-2, changes in the employment share of father's occupation (measured in percentage points and calculated for fathers' 1870 census division) between 1850-1860 and 1860-1870 are also included in the matching process (see Appendix C.1.4 for more details). In Columns 3-4, exclusively those fathers are included who worked in durable or non-durable manufacturing in 1870. In this specification, the matching process chooses a single control father owing to the reduction in the number of potential control occupations. In Columns 5-6, the matching process balances the sample on maternal characteristics (1870): a literacy and a native-born status indicator, and her age in 1870 (in years). In Columns 7-8, carpenter, truck & tractor driver and shoemaker fathers are excluded from the control group in matching. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C9: Main outcomes - fathers (G1; 1870-1900)

	Occupational change (Yes=1)		Agricultural occupation in 1900 (Yes=1)		Migration (across states) (Yes=1)		Urban in 1900 (Yes=1)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Machinist (G1)	-0.098***	-0.110***	-0.037***	-0.030***	0.038**	0.030**	0.073***	0.070***
	(0.015)	(0.016)	(0.010)	(0.009)	(0.015)	(0.012)	(0.017)	(0.016)
Unbalanced controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grandfather (G0) controls	No	Yes	No	Yes	No	Yes	No	Yes
Sample size	5429	5456	5429	5456	5429	5456	5429	5456
Number of clusters	48	49	48	49	48	49	48	49

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1900) level. All specifications are weighted by weights gained from propensity score matching described in the main text (every even column matches on grandfather controls as well). The sample includes 1120 (1116 with grandfather controls included) matched machinist fathers. The outcome variable is a binary variable which equals one if i) the father changed occupation (Col. 1-2) and the new occupation is agricultural (Col. 3-4 - farmer, farm manager/foreman/laborer), ii) he migrated across states between 1870 and 1900 (Col. 5-6) , iii) he lived in an urban place in 1900 (Col. 7-8). Unbalanced controls included in the regressions are characteristics whose mean between machinist and control fathers is still significantly different at 5% after matching. Grandfather controls are (all measured in 1860): a literacy indicator, age (measured in years), an indicator if the father lived in the same state in 1870 as the grandfather in 1860 but in a different county, an indicator if the father lived in a different state in 1870 from the grandfather in 1860, indicator variables for the grandfather holding an agricultural (farmer, farm manager/foreman/laborer) or manufacturing (durable or non-durable manufacturing) occupation, indicator variables if the grandfather worked for the railways (railroad conductor, locomotive engineer, locomotive fireman, brakeman, switchman) / in the metal industry (molder, structural metal worker, furnaceman, heater, filer, grinder, polisher, roller, tinsmith and coppersmith) / in the chemical industry (*IND1950:* Cement, concrete, gypsum and plaster products; Miscellaneous chemicals and allied products; Petroleum refining; Miscellaneous petroleum and coal products; Rubber products) / in the steel and iron industry (*IND1950:* Blast furnaces, steel works, and rolling mills; Other primary iron and steel industries; Fabricated steel products) / in machinery (*IND1950:* Agricultural machinery and tractors; Office and store machines and devices; Miscellaneous machinery; Electrical machinery, equipment, and supplies) / as a machinist, an indicator for urban status, the number of inhabitants in the place of living (*SIZEPL*), and the value of personal property and real estates. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C10: Robustness checks with regressions - sons (G2; 1900)

	Spatial sorting pre-1870		State-occupation pre-trends		Weighting		Restricted control occs	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Education rank	Urban (Yes=1)	Education rank	Urban (Yes=1)	Education rank	Urban (Yes=1)	Education rank	Urban (Yes=1)
Machinist (G1)	2.587***	0.037***	2.782***	0.035***	3.166***	0.039***	3.162***	0.039***
	(0.898)	(0.012)	(0.678)	(0.009)	(0.730)	(0.010)	(0.655)	(0.011)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County-fixed effects (1870)	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Detailed fixed effects	Yes	Yes	No	No	No	No	No	No
Sample size	55770	55770	61796	61796	63857	63857	52046	52046
Number of clusters	45	45	45	45	45	45	45	45

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1900) level. None of the specifications is weighted - except for Columns 5-6, where we use inverse proportional weights (see Appendix C.2 for details). The sample includes all sons whose father held an occupation between the 24.7th and 84.7th education rank percentiles in 1870 - except for Columns 7-8, where these cutoffs are 44.7 and 64.7, respectively. The outcome variable is the educational rank of occupation (every odd column) or a binary variable which equals one if the son lived in an urban place in 1900 (every even column). Baseline controls are described in Appendix C.1.1. Detailed fixed effects are generated by interacting the state of birth (county for the foreign-born) indicator, the county of residence indicator (1870), an urban place of living indicator (1870), and an indicator if the father was at least 34 years old in 1870. In Columns 3-4, state-occupation level measures of migration (within and across states jointly), occupation change probability, change in urban status and the probability of switching for an agricultural occupation (farmer, farm manager/foreman/laborer) are included for the 1850s and 1860s (see Appendix C.1.5 for details). Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C11: Ability bias - fathers (G1 in 1870)

	Having a child (Yes=1)		Number of children		Having a spouse (Yes=1)		Literate spouse (Yes=1)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Machinist (G1)	-0.005	0.041	-0.002	0.053	-0.005	0.018	0.003	-0.037
	(0.010)	(0.042)	(0.024)	(0.095)	(0.009)	(0.042)	(0.005)	(0.036)
Grandfather (G0)-fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
County-fixed effects (1870)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Personal controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	159716	22799	159716	22799	159716	22799	90473	9581
R^2	0.38	0.73	0.41	0.76	0.43	0.77	0.41	0.76

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are multiway clustered at the grandfather-county (1900) level. None of the specifications is weighted. The sample includes all fathers who held an occupation between the 24.7th and 84.7th education rank percentiles in 1870. In every column, the age of included fathers is between 16 and 50 years (inclusive). The outcome variable is i) a binary variable which equals one if the father had at least one child in 1870 (Col. 1-2), ii) the number of children in 1870 (Col. 3-4); iii)-iv) a binary variable which equals one if the father had a spouse (Col. 5-6) and, conditional on having a wife, she was literate (Col. 7-8). Personal controls included in the regressions are (all measured in 1870): the education rank of occupation, urban status and literacy indicator, age (in years), value of real estate and personal property, number of inhabitants in the place of living and a farmer-farm manager-farm foreman indicator. The interactions of the urban indicator, size of place of living, two wealth measures, education rank and age are also included. The squared size of place of living, wealth measures and age are included as well. Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C12: Robustness checks by the age of fathers and sons - sons (G2 in 1900)

	Sons (0-5 y.o.) of young fathers (<33 y.o.; G1)		Sons (0-5 y.o.) of old fathers (>=33 y.o.; G1)		Old and young sons
	(1)	(2)	(3)	(4)	(5)
	Education rank	Urban (Yes=1)	Education rank	Urban (Yes=1)	Education rank
Machinist (G1)	4.176***	0.041**	3.616***	0.044**	4.001***
	(1.397)	(0.020)	(1.106)	(0.020)	(0.695)
Machinist (G1) x					-4.098***
$\mathbb{1}$(son (G2) older than 10 y.o. in 1870)					(1.323)
Baseline controls	Yes	Yes	Yes	Yes	Yes
County-fixed effects (1870)	Yes	Yes	Yes	Yes	Yes
Sample size	16338	16338	18266	18266	63857
Number of clusters	45	45	45	45	45

Note: OLS regression coefficients with standard errors in parentheses. Standard errors are clustered at the state (1900) level. None of the specifications is weighted. The sample includes all sons i) whose father held an occupation between the 24.7th and 84.7th education rank percentiles in 1870; and ii) who were not older than five years in 1870 (Col. 1-4). The outcome variable is the educational rank of occupation (Col. 1,3,5) or a binary variable which equals one if the son lived in an urban place in 1900 (Col. 2,4). Column 5 includes the son age indicator as a main effect separately. Baseline controls are described in Appendix C.1.1. The estimation includes only fathers who were younger (older) than thirty-three years in 1870 in Columns 1-2 (3-4). Levels of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Bibliography

Abramitzky, Ran, Philipp Ager, Leah Platt Boustan, Elior Cohen, and Casper Worm Hansen (2019a). *The Effects of Immigration on the Economy: Lessons from the 1920s Border Closure*. NBER Working Papers 26536. National Bureau of Economic Research, Inc.

Abramitzky, Ran, Leah Boustan, Elisa Jacome, and Santiago Perez (2021). «Intergenerational Mobility of Immigrants in the United States over Two Centuries». In: *American Economic Review* 111.2, pp. 580–608.

Abramitzky, Ran, Leah Boustan, and Myera Rashid (2020). *Census Linking Project: Version 1.0 [dataset]. 2020.* URL: https://censuslinkingproject.org.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson (2012). «Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration». In: *American Economic Review* 102.5, pp. 1832–1856.

— (2013). «Have the poor always been less likely to migrate? Evidence from inheritance practices during the age of mass migration». In: *Journal of Development Economics* 102.C, pp. 2–14.

— (2019b). «To the New World and Back Again: Return Migrants in the Age of Mass Migration». In: *ILR Review* 72.2, pp. 300–322.

Acemoglu, Daron (1998). «Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality». In: *The Quarterly Journal of Economics* 113.4, pp. 1055–1089.

Acemoglu, Daron and David Autor (2011). «Skills, Tasks and Technologies: Implications for Employment and Earnings». In: *Handbook of Labor Economics*. Ed. by O. Ashenfelter and D. Card. Vol. 4. Handbook of Labor Economics. Elsevier. Chap. 12, pp. 1043–1171.

Acemoglu, Daron and Pascual Restrepo (2019). «Automation and New Tasks: How Technology Displaces and Reinstates Labor». In: *Journal of Economic Perspectives* 33.2, pp. 3–30.

— (2020). «Robots and Jobs: Evidence from US Labor Markets». In: *Journal of Political Economy* 128.6, pp. 2188–2244.

Ager, Philipp and Casper Worm Hansen (2017). *Closing Heaven's Door: Evidence from the 1920s U.S.Immigration Quota Acts*. Discussion Papers 17-22. University of Copenhagen. Department of Economics.

Ager, Philipp, Benedikt Herz, and Markus Brueckner (2020a). «Structural Change and the Fertility Transition». In: *The Review of Economics and Statistics* 102.4, pp. 806–822.

— (2020b). «Structural Change and the Fertility Transition». In: *The Review of Economics and Statistics* 102.4, pp. 806–822.

Aigner, Dennis J. (1973). «Regression with a binary independent variable subject to errors of observation». In: *Journal of Econometrics* 1.1, pp. 49–59.

Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney (2016). «The Long-Run Impact of Cash Transfers to Poor Families». In: *American Economic Review* 106.4, pp. 935–971.

Akee, Randall K. Q., William E. Copeland, Gordon Keeler, Adrian Angold, and E. Jane Costello (2010). «Parents' Incomes and Children's Outcomes: A Quasi-experiment Using Transfer Payments from Casino Profits». In: *American Economic Journal: Applied Economics* 2.1, pp. 86–115.

Alston, Lee J. and Timothy J. Hatton (1991). «The Earnings Gap Between Agricultural and Manufacturing Laborers, 1925-1941». In: *The Journal of Economic History* 51.1, pp. 83–99.

Alvarez-Cuadrado, Francisco and Markus Poschke (2011). «Structural Change Out of Agriculture: Labor Push versus Labor Pull». In: *American Economic Journal: Macroeconomics* 3.3, pp. 127–158.

Andersson, David, Mounir Karadja, and Erik Prawitz (2020). «Mass Migration and Technological Change». Unpublished manuscript.

Angelucci, Manuela (2015). «Migration and Financial Constraints: Evidence from Mexico». In: *The Review of Economics and Statistics* 97.1, pp. 224–228.

Antman, Francisca M. (2011). «The intergenerational effects of paternal migration on schooling and work: What can we learn from children's time allocations?» In: *Journal of Development Economics* 96.2, pp. 200–208.

Arkolakis, Costas, Michael Peters, and Sun Kyoung Lee (2019). *European Immigrants and the United States' Rise to the Technological Frontier*. 2019 Meeting Papers 1420. Society for Economic Dynamics.

Artuç, Erhan, Shubham Chaudhuri, and John McLaren (2010). «Trade Shocks and Labor Adjustment: A Structural Empirical Approach». In: *American Economic Review* 100.3, pp. 1008–45.

Ashenfelter, Orley and Cecilia Rouse (1998). «Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins». In: *The Quarterly Journal of Economics* 113.1, pp. 253–284.

Atack, Jeremy, Fred Bateman, and Robert A. Margo (2004). «Skill Intensity and Rising Wage Dispersion in Nineteenth-Century American Manufacturing». In: *The Journal of Economic History* 64.1, pp. 172–192.

Atack, Jeremy, Robert A. Margo, and Paul Rhode (2020). *'Mechanization Takes Command': Inanimate Power and Labor Productivity in Late Nineteenth Century American Manufacturing*. NBER Working Papers 27436. National Bureau of Economic Research, Inc.

Atkin, David (2016). «Endogenous Skill Acquisition and Export Manufacturing in Mexico». In: *American Economic Review* 106.8, pp. 2046–2085.

Austin, Peter C. (2011). «An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies». In: *Multivariate Behavioral Research* 46.3, pp. 399–424.

Autor, David H., David Dorn, and Gordon H. Hanson (2016). «The China Shock: Learning from Labor-Market Adjustment to Large Changes in Trade». In: *Annual Review of Economics* 8.1, pp. 205–240.

Bailey, Martha J., Connor Cole, Morgan Henderson, and Catherine Massey (2020). «How Well Do Automated Linking Methods Perform? Lessons from US Historical Data». In: *Journal of Economic Literature* 58.4, pp. 997–1044.

Bartelme, Dominick G., Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare (2019). *The Textbook Case for Industrial Policy: Theory Meets Data*. NBER Working Papers 26193. National Bureau of Economic Research, Inc.

Baten, Jörg, Anna Spadavecchia, Jochen Streb, and Shuxi Yin (2007). «What made southwest German firms innovative around 1900? Assessing the importance of intra- and inter-industry externalities». In: *Oxford Economic Papers* 59, pp. 105–126.

Bazzi, Samuel (2017). «Wealth Heterogeneity and the Income Elasticity of Migration». In: *American Economic Journal: Applied Economics* 9.2, pp. 219–255.

Beaudry, Paul and David A. Green (2003). «Wages and Employment in the United States and Germany: What Explains the Differences?» In: *American Economic Review* 93.3, pp. 573–602.

— (2005). «Changes in U.S. Wages, 1976-2000: Ongoing Skill Bias or Major Technological Change?» In: *Journal of Labor Economics* 23.3, pp. 609–648.

Becker, Gary and Nigel Tomes (1979). «An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility». In: *Journal of Political Economy* 87.6, pp. 1153–89.

— (1986). «Human Capital and the Rise and Fall of Families». In: *Journal of Labor Economics* 4.3, S1–39.

Beine, Michel, Fréderic Docquier, and Hillel Rapoport (2008). «Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers». In: *Economic Journal* 118.528, pp. 631–652.

Bengtsson, Erik (2012). «Labor's share in Sweden, 1850-2000». Unpublished manuscript.

Bernáth, Dezső (1902). «Belső munkás-vándorlások Magyarországon». In: *Közgazdasági Szemle* 28, pp. 582–592.

Bettinger, Eric, Oded Gurantz, Laura Kawano, Bruce Sacerdote, and Michael Stevens (2019). «The Long-Run Impacts of Financial Aid: Evidence from California's Cal Grant». In: *American Economic Journal: Economic Policy* 11.1, pp. 64–94.

Bingley, Paul and Alessandro Martinello (2017). «Measurement Error in Income and Schooling and the Bias of Linear Estimators». In: *Journal of Labor Economics* 35.4, pp. 1117–1148.

Binswanger, Hans P. (1974). «A Cost Function Approach to the Measurement of Elasticities of Factor Demand and Elasticities of Substitution». In: *American Journal of Agricultural Economics* 56.2, pp. 377–386.

Binzel, Christine and Ragui Assaad (2011). «Egyptian men working abroad: Labour supply responses by the women left behind». In: *Labour Economics* 18.S1, pp. 98–114.

Bleakley, Hoyt and Joseph Ferrie (2016). «Shocking Behavior: Random Wealth in Antebellum Georgia and Human Capital Across Generations». In: *The Quarterly Journal of Economics* 131.3, pp. 1455–1495.

Bodovics, Éva (2019). «A amerikai kivándorlás éghajlattörténeti aspektusai». In: *Amerikai magyarok - magyar amerikaiak*. Ed. by László Ambrus and Eszter Rakita. Líceum Kiadó.

Bolt, Jutta and Jan Luiten van Zanden (2020). «Maddison style estimates of the evolution of the world economy. A new 2020 update». In: *Maddison Project Working Paper*.

Boneva, Teodora and Christopher Rauh (2018). «Parental Beliefs about Returns to Educational Investments—The Later the Better?» In: *Journal of the European Economic Association* 16.6, pp. 1669–1711.

Boone, Christopher and Laurence Wilse-Samson (2019). «Farm Mechanization and Rural Migration in the Great Depression». Unpublished manuscript.

Boppart, Timo (2014). «Structural Change and the Kaldor Facts in a Growth Model With Relative Price Effects and Non-Gorman Preferences». In: *Econometrica* 82, pp. 2167–2196.

Boserup, Simon Halphen, Wojciech Kopczuk, and Claus Thustrup Kreiner (2018). «Born with a Silver Spoon? Danish Evidence on Wealth Inequality in Childhood». In: *Economic Journal* 128.612, pp. 514–544.

Bratberg, Espen, Øivind Anti Nilsen, and Kjell Vaage (2008). «Job losses and child outcomes». In: *Labour Economics* 15.4, pp. 591–603.

Buggle, Johannes C. and Steven Nafziger (2021). «The Slow Road from Serfdom: Labor Coercion and Long-Run Development in the Former Russian Empire». In: *The Review of Economics and Statistics* 103.1, pp. 1–17.

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli (2016). «Agricultural Productivity and Structural Transformation: Evidence from Brazil». In: *American Economic Review* 106.6, pp. 1320–1365.

Bustos, Paula, Juan Manuel Castro-Vincenzi, Joan Monras, and Jacopo Ponticelli (2019). *Industrialization without Innovation*. NBER Working Papers 25871. National Bureau of Economic Research, Inc.

Card, David and Alan B Krueger (1992). «Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States». In: *Journal of Political Economy* 100.1, pp. 1–40.

Caselli, Francesco and Wilbur John II Coleman (2006). «The World Technology Frontier». In: *American Economic Review* 96.3, pp. 499–522.

Castleman, Benjamin L. and Bridget Terry Long (2016). «Looking beyond Enrollment: The Causal Effect of Need-Based Grants on College Access, Persistence, and Graduation». In: *Journal of Labor Economics* 34.4, pp. 1023–1073.

Caucutt, Elizabeth M. and Lance Lochner (2020). «Early and Late Human Capital Investments, Borrowing Constraints, and the Family». In: *Journal of Political Economy* 128.3, pp. 1065–1147.

Cesarini, David, Erik Lindqvist, Matthew J. Notowidigdo, and Robert Östling (2017). «The Effect of Wealth on Individual and Household Labor Supply: Evidence from Swedish Lotteries». In: *American Economic Review* 107.12, pp. 3917–3946.

Chandler, Alfred D. Jr. (1977). *The Visible Hand: The Managerial Revolution in American Business*. The Belknap Press.

— (1994). *Scale and Scope: The Dynamics of Industrial Capitalism*. The Belknap Press.

Chernina, Eugenia, Paul Castañeda Dower, and Andrei Markevich (2014). «Property rights, land liquidity, and internal migration». In: *Journal of Development Economics* 110.C, pp. 191–215.

Chetty, Raj and Nathaniel Hendren (2018a). «The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects». In: *The Quarterly Journal of Economics* 133.3, pp. 1107–1162.

— (2018b). «The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates». In: *The Quarterly Journal of Economics* 133.3, pp. 1163–1228.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez (2014). «Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States». In: *The Quarterly Journal of Economics* 129.4, pp. 1553–1623.

Ciccone, Antonio (2002). «Input Chains and Industrialization». In: *Review of Economic Studies* 69.3, pp. 565–587.

Cinelli, Carlos and Chad Hazlett (2020). «Making sense of sensitivity: extending omitted variable bias». In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82.1, pp. 39–67.

Clemens, Michael (2020). *The Emigration Life Cycle: How Development Shapes Emigration from Poor Countries*. IZA Discussion Papers 13614. Institute of Labor Economics (IZA).

Clemens, Michael, Ethan Lewis, and Hannah Postel (2018). «Immigration Restrictions as Active Labor Market Policy: Evidence from the Mexican Bracero Exclusion». In: *American Economic Review* 108.6, pp. 1468–1487.

Coelli, Michael B. (2011). «Parental job loss and the education enrollment of youth». In: *Labour Economics* 18.1, pp. 25–35.

Collins, William J. and Marianne H. Wanamaker (2014). «Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data». In: *American Economic Journal: Applied Economics* 6.1, pp. 220–252.

— (2017). *African American Intergenerational Economic Mobility Since 1880*. NBER Working Papers 23395. National Bureau of Economic Research, Inc.

Combes, Pierre-Philippe and Laurent Gobillon (2015). «The Empirics of Agglomeration Economies». In: *Handbook of Regional and Urban Economics*. Ed. by Gilles Duranton, Vernon Henderson, and William C. Strange. Vol. 5. Handbook of Regional and Urban Economics. Elsevier. Chap. 0, pp. 247–348.

Combes, Pierre-Philippe, Miren Lafourcade, Jacques-François Thisse, and Jean-Claude Toutain (2011). «The rise and fall of spatial inequalities in France: A long-run perspective». In: *Explorations in Economic History* 48.2, pp. 243–271.

Comin, Diego and Bart Hobijn (2004). «Cross-country technology adoption: making the theories face the facts». In: *Journal of Monetary Economics* 51.1, pp. 39–83.

Comin, Diego, Bart Hobijn, and Emilie Rovito (2008). «Technology usage lags». In: *Journal of Economic Growth* 13.4, pp. 237–256.

Conley, Timothy G. (1999). «GMM estimation with cross sectional dependence». In: *Journal of Econometrics* 92.1, pp. 1–45.

Cooper, Kerris and Kitty Stewart (2017). *Does Money Affect Children's Outcomes? An update*. CASE Papers 203. Centre for Analysis of Social Exclusion, LSE.

Correia, Sergio (2016). *Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator*. Tech. rep. Working Paper.

Cortes, Guido Matias and Giovanni Gallipoli (2018). «The Costs of Occupational Mobility: An Aggregate Analysis». In: *Journal of the European Economic Association* 16.2, pp. 275–315.

Crafts, Nicholas and Abay Mulatu (2006). «How Did the Location of Industry Respond to Falling Transport Costs in Britain Before World War I?» In: *The Journal of Economic History* 66.3, pp. 575–607.

Cvrcek, Tomas (2013). «Wages, Prices, and Living Standards in the Habsburg Empire, 1827-1910». In: *The Journal of Economic History* 73.1, pp. 1–37.

Dahl, Gordon B. and Lance Lochner (2012). «The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit». In: *American Economic Review* 102.5, pp. 1927–1956.

Dao, Thu Hien, Frédéric Docquier, Chris Parsons, and Giovanni Peri (2018). «Migration and development: Dissecting the anatomy of the mobility transition». In: *Journal of Development Economics* 132.C, pp. 88–101.

Dauth, Wolfgang, Sebastian Findeisen, and Jens Suedekum (2021a). «Adjusting to Globalization in Germany». In: *Journal of Labor Economics* 39.1, pp. 263–302.

Dauth, Wolfgang, Sebastian Findeisen, Jens Suedekum, and Nicole Woessner (2021b). «The Adjustment of Labor Markets to Robots». In: *Journal of the European Economic Association*.

David, Anda and Joachim Jarreau (2017). «Migrating out of a crowded labor market: evidence from Egypt». In: *IZA Journal of Migration and Development* 7.1, pp. 1–28.

David, Paul A. (1990). «The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox». In: *American Economic Review* 80.2, pp. 355–361.

Dawson, Andrew (1979). «The paradox of dynamic technological change and the labor aristocracy in the United States, 1880–1914». In: *Labor History* 20.3, pp. 325–351.

Denning, Jeffrey T., Benjamin M. Marx, and Lesley J. Turner (2019). «ProPelled: The Effects of Grants on Graduation, Earnings, and Welfare». In: *American Economic Journal: Applied Economics* 11.3, pp. 193–224.

De Pleijt, Alexandra, Alessandro Nuvolari, and Jacob Weisdorf (2020). «Human Capital Formation During the First Industrial Revolution: Evidence from the use of Steam Engines». In: *Journal of the European Economic Association* 18.2, pp. 829–889.

De Pleijt, Alexandra and Jacob Weisdorf (2017). «Human capital formation from occupations: the 'deskilling hypothesis' revisited». In: *Cliometrica* 11.1, pp. 1–30.

Di Maio, Michele and Roberto Nisticò (2019). «The effect of parental job loss on child school dropout: Evidence from the Occupied Palestinian Territories». In: *Journal of Development Economics* 141.C.

Diamond, Rebecca (2016). «The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000». In: *American Economic Review* 106.3, pp. 479–524.

Dinopoulos, Elias and Laixun Zhao (2007). «Child Labor and Globalization». In: *Journal of Labor Economics* 25, pp. 553–579.
Diodato, Dario, Frank Neffke, and Neave O'Clery (2018). «Why do industries coagglomerate? How Marshallian externalities differ by industry and have evolved over time». In: *Journal of Urban Economics* 106.C, pp. 1–26.
Dix, Keith (1988). *What's a Coal Miner to Do? The Mechanization of Coal Mining*. University of Pittsburgh Press.
Dix-Carneiro, Rafael (2014). «Trade Liberalization and Labor Market Dynamics». In: *Econometrica* 82.3, pp. 825–885.
Dobbie, Will and Roland G. Fryer (2011). «Are High-Quality Schools Enough to Increase Achievement among the Poor? Evidence from the Harlem Children's Zone». In: *American Economic Journal: Applied Economics* 3.3, pp. 158–187.
Docquier, Frédéric and Hillel Rapoport (2012). «Globalization, Brain Drain, and Development». In: *Journal of Economic Literature* 50.3, pp. 681–730.
Doran, Kirk and Chungeun Yoon (2020). «Immigration and Invention: Evidence from the Quota Act». Unpublished manuscript.
Dornbusch, Rudiger, Stanley Fischer, and Paul A Samuelson (1977). «Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods». In: *American Economic Review* 67.5, pp. 823–839.
Duflo, Esther (2001). «Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment». In: *American Economic Review* 91.4, pp. 795–813.
Dupraz, Yannick and Andreas Ferrara (2021). *Fatherless: The Long-Term Effects of Losing a Father in the U.S. Civil War*. CAGE Online Working Paper Series 538. Competitive Advantage in the Global Economy (CAGE).
Duranton, Gilles and Diego Puga (2001). «Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products». In: *American Economic Review* 91.5, pp. 1454–1477.
Durlauf, Steven (2004). «Neighborhood effects». In: *Handbook of Regional and Urban Economics*. Ed. by J. V. Henderson and J. F. Thisse. 1st ed. Vol. 4. Elsevier. Chap. 50, pp. 2173–2242.
Dustmann, Christian, Itzhak Fadlon, and Yoram Weiss (2011). «Return migration, human capital accumulation and the brain drain». In: *Journal of Development Economics* 95.1, pp. 58–67.
Dustmann, Christian and Albrecht Glitz (2015). «How Do Industries and Firms Respond to Changes in Local Labor Supply?» In: *Journal of Labor Economics* 33.3, pp. 711–750.
Dustmann, Christian, Albrecht Glitz, and Tommaso Frattini (2008). «The labour market impact of immigration». In: *Oxford Review of Economic Policy* 24.3, pp. 478–495.

Dustmann, Christian and Anna Okatenko (2014). «Out-migration, wealth constraints, and the quality of local amenities». In: *Journal of Development Economics* 110.C, pp. 52–63.

Éber, Ernő (1902). *Munkabér és Kivándorlás*. Pesti Nyomda.

Eddie, Scott M. (1968). «Agricultural production and output per worker in Hungary, 1870-1913». In: *The Journal of Economic History* 28.2, pp. 197–222.

Edin, Per-Anders, Tiernan Evans, Georg Graetz, Sofia Hernnäs, and Guy Michaels (2019). *Individual Consequences of Occupational Decline*. CEPR Discussion Papers 13808. C.E.P.R. Discussion Papers.

Ellison, Glenn and Edward L Glaeser (1997). «Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach». In: *Journal of Political Economy* 105.5, pp. 889–927.

Emerick, Kyle (2018). «Agricultural productivity and the sectoral reallocation of labor in rural India». In: *Journal of Development Economics* 135, pp. 488–503.

Engerman, Stanley and Claudia Goldin (1991). *Seasonality in Nineteenth Century Labor Markets*. NBER Historical Working Papers 0020. National Bureau of Economic Research, Inc.

Erdélyi Riport (2014). *A Monarchia jónéven vette a kivándorlást*. http://erdelyiriport.ro/interju/a-monarchia-joneven-vette-a-kivandorlast, Accessed: 09/12/2019.

Eriksson, Katherine and Zachary Ward (2019). «The Residential Segregation of Immigrants in the United States from 1850 to 1940». In: *The Journal of Economic History* 79.4, pp. 989–1026.

Fack, Gabrielle and Julien Grenet (2015). «Improving College Access and Success for Low-Income Students: Evidence from a Large Need-Based Grant Program». In: *American Economic Journal: Applied Economics* 7.2, pp. 1–34.

Federico, Giovanni and Alexander Klein (2010). «Sectoral developments, 1870-1914». In: *The Cambridge Economic History of Modern Europe*. Ed. by Stephen Broadberry and Kevin H. O'Rourke. Vol. 2. Cambridge University Press, pp. 59–83.

Feigenbaum, James and Daniel P. Gross (2020). *Automation and the Fate of Young Workers: Evidence from Telephone Operation in the Early 20th Century*. NBER Working Papers 28061. National Bureau of Economic Research, Inc.

Feigenbaum, James and Hui Ren Tan (2020). «The Return to Education in the Mid-Twentieth Century: Evidence from Twins». In: *The Journal of Economic History* 80.4, pp. 1101–1142.

Fenyvessy, József (1902). «Iparfelügyelet». In: *Közgazdasági Szemle* 28, pp. 551–555.

Fernández-Sánchez, Martín (2020). «Mass Emigration and Human Capital over a Century: Evidence from the Galician Diaspora». Unpublished manuscript.

Franck, Raphaël and Oded Galor (2021). «Flowers of evil? Industrialization and long run development». In: *Journal of Monetary Economics* 117.C, pp. 108–128.

Frank, Alison Fleig (2007). *Oil Empire*. Harvard University Press.

Frisnyák, Zsuzsanna (2006). «Áruforgalom, áruszállítás és a magyarországi vasutak (1895). Vasút–ember–tér kapcsolat». Eötvös Loránd Tudományegyetem Bölcsészettudományi Kar.

Fuchs, Simon (2018). «The Spoils of War: Trade Shocks during WWI and Spain's Regional Development». Unpublished manuscript.

Gabbuti, Giacomo (2018). *Labour Shares and Income Inequality: Insights from Italian Economic History, 1895-2015*. HHB Working Papers Series 13. The Historical Household Budgets Project.

Galster, George C. (2012). «The Mechanism(s) of Neighbourhood Effects: Theory, Evidence, and Policy Implications». In: *Neighbourhood Effects Research: New Perspectives*. Ed. by Maarten van Ham, David Manley, Nick Bailey, Ludi Simpson, and Duncan Maclennan. Springer Netherlands.

Gao, Pei and Yu-Hsiang Lei (2021). «Communication Infrastructure and Stabilizing Food Prices: Evidence from the Telegraph Network in China». In: *American Economic Journal: Applied Economics* 13.3, pp. 65–101.

Gervais, Antoine and J. Bradford Jensen (2019). «The tradability of services: Geographic concentration and trade costs». In: *Journal of International Economics* 118.C, pp. 331–350.

Gibson, Campbell and Emily Lennon (1999). *Historical Census Statistics on the Foreign-Born Population of the United States: 1850 to 1990*. Working Paper 29. US Census Bureau.

Gibson, John and David McKenzie (2014). «Development through seasonal worker programs: the case of New Zealand's RSE program». In: *International Handbook on Migration and Economic Development*. Ed. by Robert E.B. Lucas. Chapters. Edward Elgar Publishing. Chap. 7, pp. 186–210.

Goldin, Claudia (1998). «America's Graduation from High School: The Evolution and Spread of Secondary Schooling in the Twentieth Century». In: *The Journal of Economic History* 58.2, pp. 345–374.

Goldin, Claudia and Lawrence F. Katz (1998). «The Origins of Technology-Skill Complementarity». In: *The Quarterly Journal of Economics* 113.3, pp. 693–732.

— (1999). «Human Capital and Social Capital: The Rise of Secondary Schooling in America, 1910-1940». In: *The Journal of Interdisciplinary History* 29.4, pp. 683–723.

— (2000). «Education and Income in the Early Twentieth Century: Evidence from the Prairies». In: *The Journal of Economic History* 60.3, pp. 782–818.

— (2008). *The Race Between Education and Technology.* Belknap Press for Harvard University Press.

Goldin, Claudia and Robert A. Margo (1992). «The Great Compression: The Wage Structure in the United States at Mid-Century». In: *The Quarterly Journal of Economics* 107.1, pp. 1–34.

Gollin, Douglas, Stephen Parente, and Richard Rogerson (2002). «The Role of Agriculture in Development». In: *American Economic Review* 92.2, pp. 160–164.

Gollin, Douglas, Stephen L. Parente, and Richard Rogerson (2007). «The food problem and the evolution of international income levels». In: *Journal of Monetary Economics* 54.4, pp. 1230–1255.

González, Libertad and Francesc Ortega (2011). «How do very open economies adjust to large immigration flows? Evidence from Spanish regions». In: *Labour Economics* 18.1, pp. 57–70.

Good, David F. (1984). *The Economic Rise of the Habsburg Empire: 1750-1914*. University of California Press.

Graetz, Georg and Guy Michaels (2018). «Robots at Work». In: *The Review of Economics and Statistics* 100.5, pp. 753–768.

Gray, Rowena (2013). «Taking technology to task: The skill content of technological change in early twentieth century United States». In: *Explorations in Economic History* 50.3, pp. 351–367.

Gray, Rowena, Gaia Narciso, and Gaspare Tortorici (2019). «Globalization, agricultural markets and mass migration: Italy, 1881–1912». In: *Explorations in Economic History* 74, p. 101276.

Guriev, Sergei and Elena Vakulenko (2015). «Breaking out of poverty traps: Internal migration and interregional convergence in Russia». In: *Journal of Comparative Economics* 43.3, pp. 633–649.

Hai, Rong and James J. Heckman (2017). «Inequality in Human Capital and Endogenous Credit Constraints». In: *Review of Economic Dynamics* 25, pp. 4–36.

Haines, Michael R. (1989). «A State and Local Consumer Price Index for the United States in 1890». In: *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 22.3, pp. 97–105.

Hanlon, W. Walker (2015). «Necessity Is the Mother of Invention: Input Supplies and Directed Technical Change». In: *Econometrica* 83, pp. 67–100.

Hanlon, W. Walker and Antonio Miscio (2017). «Agglomeration: A long-run panel data approach». In: *Journal of Urban Economics* 99.C, pp. 1–14.

Hanlon, Walker (2021). «The Rise of the Engineer: Inventing the Professional Inventor During the Industrial Revolution». Unpublished manuscript.

Hanson, Gordon H. and Matthew J. Slaughter (2002). «Labor-market adjustment in open economies: Evidence from US states». In: *Journal of International Economics* 57.1, pp. 3–29.

Hatton, Timothy J. (1995). «A model of Scandinavian emigration, 1870-1913». In: *European Economic Review* 39.3-4, pp. 557–564.

Hatton, Timothy J. and Jeffrey G. Williamson (1991). «Wage gaps between farm and city: Michigan in the 1890s». In: *Explorations in Economic History* 28.4, pp. 381–408.

— (1994). «What Drove the Mass Migrations from Europe in the Late Nineteenth Century?» In: *Population and Development Review* 20.3, pp. 533–559.

Heckman, James J. and Flavio Cunha (2007). «The Technology of Skill Formation». In: *American Economic Review* 97.2, pp. 31–47.

Hegedüs, Loránt (1899). *A magyarok kivándorlása Amerikába*. Budapest.

— (1911). «Az ipari hullám és kivándorlási konjunktúra». In: *Magyar Gyáripar* 1.17, pp. 3–4.

Henderson, Vernon, Todd Lee, and Yung Joon Lee (2001). «Scale Externalities in Korea». In: *Journal of Urban Economics* 49.3, pp. 479–504.

Henderson, Vernon and Matthew A. Turner (2020). «Urbanization in the Developing World: Too Early or Too Slow?» In: *Journal of Economic Perspectives* 34.3, pp. 150–73.

Hilger, Nathaniel G. (2016). «Parental Job Loss and Children's Long-Term Outcomes: Evidence from 7 Million Fathers' Layoffs». In: *American Economic Journal: Applied Economics* 8.3, pp. 247–83.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart (2007). «Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference». In: *Political Analysis* 15.3, pp. 199–236.

Hornbeck, Richard and Suresh Naidu (2014). «When the Levee Breaks: Black Migration and Economic Development in the American South». In: *American Economic Review* 104.3, pp. 963–990.

Humlum, Anders (2019). «Robot Adoption and Labor Market Dynamics». Unpublished manuscript.

Hungarian Royal Central Statistical Office (1918). «Emigration and return migration in the countries of the Holy Crown of Hungary (1899-1913)». In: *Hungarian Statistical Bulletin - New Series* 67.

Imai, Kosuke, Gary King, and Elizabeth A. Stuart (2008). «Misunderstandings between experimentalists and observationalists about causal inference». In: *Journal of the Royal Statistical Society Series A* 171.2, pp. 481–502.

Imbert, Clément, Marlon Seror, Yanos Zylberberg, and Yifan Zhang (2020). *Migrants and Firms : Evidence from China*. CAGE Online Working Paper Series 460. Competitive Advantage in the Global Economy (CAGE).

Inwood, Kris, Chris Minns, and Fraser Summerfield (2019). «Occupational income scores and immigrant assimilation. Evidence from the Canadian census». In: *Explorations in Economic History* 72.C, pp. 114–122.

Jerome, Harry (1926). *Migration and Business Cycles*. NBER Books. National Bureau of Economic Research, Inc.

— (1934). *Mechanization in Industry*. NBER Books. National Bureau of Economic Research, Inc.

Jones, Ronald W. (1965). «The Structure of Simple General Equilibrium Models». In: *Journal of Political Economy* 73.6, pp. 557–572.

Kaboski, Joseph P. and Trevon D. Logan (2011). «Factor Endowments and the Returns to Skill: New Evidence from the American Past». In: *Journal of Human Capital* 5.2, pp. 111–152.

Kambourov, Gueorgui and Iourii Manovskii (2009). «Occupational Specificity Of Human Capital». In: *International Economic Review* 50.1, pp. 63–115.

Kaposi, Zoltán (2002). *Magyar gazdaságtörténet 1700-2000*. Dialóg Campus Kiadó, Budapest-Pécs.

Kaps, Klemens (2015). *Ungleiche Entwicklung in Zentraleuropa. Galizien zwischen überregionaler Arbeitsteilung und imperialer Politik (1772-1914)*. Böhlau.

Karadja, Mounir and Erik Prawitz (2019). «Exit, Voice and Political Change: Evidence from Swedish Mass Migration to the United States». In: *Journal of Political Economy* 127.4.

Katz, Lawrence F. and Robert A. Margo (2014a). «Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective». In: *Human Capital in History: The American Record*. NBER Chapters. National Bureau of Economic Research, Inc, pp. 15–57.

— (2014b). «Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective». In: *Human Capital in History: The American Record*. NBER Chapters. National Bureau of Economic Research, Inc, pp. 15–57.

Kelly, Morgan (2020). *Understanding Persistence*. CEPR Discussion Papers 15246. C.E.P.R. Discussion Papers.

Kelly, Morgan, Joel Mokyr, and Cormac Ó Gráda (2014). «Precocious Albion: A New Interpretation of the British Industrial Revolution». In: *Annual Review of Economics* 6.1, pp. 363–389.

Kelly, Morgan, Joel Mokyr, and Cormac Ó Gráda (2020). *The Mechanics of the Industrial Revolution*. CEPR Discussion Papers 14884. C.E.P.R. Discussion Papers.

Kim, Sukkoo (1995). «Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U. S. Regional Manufacturing Structure, 1860–1987». In: *The Quarterly Journal of Economics* 110.4, pp. 881–908.

Kim, Sukkoo (1999). «Regions, resources, and economic geography: Sources of U.S. regional comparative advantage, 1880–1987». In: *Regional Science and Urban Economics* 29.1, pp. 1–32.
— (2006). «Division of labor and the rise of cities: evidence from US industrialization, 1850–1880». In: *Journal of Economic Geography* 6.4, pp. 469–491.
— (2007). *Immigration, Industrial Revolution and Urban Growth in the United States, 1820-1920: Factor Endowments, Technology and Geography*. NBER Working Papers 12900. National Bureau of Economic Research, Inc.
King, Gary and Richard Nielsen (2019). «Why Propensity Scores Should Not Be Used for Matching». In: *Political Analysis* 27.4, pp. 435–454.
Klein, Alexander and Nicholas Crafts (2020a). «Agglomeration externalities and productivity growth: US cities, 1880–1930». In: *Economic History Review* 73.1, pp. 209–232.
— (2020b). «Agglomeration externalities and productivity growth: US cities, 1880–1930». In: *Economic History Review* 73.1, pp. 209–232.
Klein, Alexander, Max-Stephan Schulze, and Tamás Vonyó (2017). «How Peripheral was the Periphery?» In: *The Spread of Modern Industry to the Periphery since 1871*. Ed. by Kevin Hjortshøj O'Rourke and Jeffrey Gale Williamson. Oxford University Press.
Kline, Patrick and Enrico Moretti (2014). «Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority». In: *The Quarterly Journal of Economics* 129.1, pp. 275–331.
Knudsen, Anne Sofie Beck (2019). *Those Who Stayed: Individualism, Self-Selection and Cultural Change during the Age of Mass Migration*. Discussion Papers 19-01. University of Copenhagen. Department of Economics.
Koffsky, Nathan (1949). «Farm and Urban Purchasing Power». In: *Studies in Income and Wealth, Volume 11*. NBER, pp. 151–220.
Kongsamut, Piyabha, Sergio Rebelo, and Danyang Xie (2001). «Beyond Balanced Growth». In: *Review of Economic Studies* 68.4, pp. 869–882.
Kozári, Mónika (2009). «A dualizmus kora». In: *Magyarország története*. Ed. by Ignác Romsics. Kossuth Kiadó. Chap. 16.
Krugman, Paul (1980). «Scale Economies, Product Differentiation, and the Pattern of Trade». In: *American Economic Review* 70.5, pp. 950–959.
— (1991). «Increasing Returns and Economic Geography». In: *Journal of Political Economy* 99.3, pp. 483–499.
Krusell, Per, Lee E. Ohanian, José-Victor Rios-Rull, and Giovanni L. Violante (2000). «Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis». In: *Econometrica* 68.5, pp. 1029–1054.

Kulcsár, László and J. László Kulcsár (2019). «”The boarding house is the life for me” Regional and socioeconomic dynamics of emigration to America before the Great War». In: *Tér és Társadalom* 33.2, pp. 83–107.

Lafortune, Jeanne, Ethan Lewis, and José Tessada (2019). «People and Machines: A Look at the Evolving Relationship between Capital and Skill in Manufacturing, 1860–1930, Using Immigration Shocks». In: *The Review of Economics and Statistics* 101.1, pp. 30–43.

Lafortune, Jeanne, Ethan G. Lewis, José Pablo Martínez, and José Tessada (2021). *Changing Returns to Scale in Manufacturing 1880-1930: The Rise of (Skilled) Labor?* NBER Working Papers 28633. National Bureau of Economic Research, Inc.

Lafortune, Jeanne, José Tessada, and Carolina González-Velosa (2015). «More hands, more power? Estimating the impact of immigration on output and technology choices using early 20th century US agriculture». In: *Journal of International Economics* 97.2, pp. 339–358.

Laliberté, Jean-William (2021). «Long-Term Contextual Effects in Education: Schools and Neighborhoods». In: *American Economic Journal: Economic Policy* 13.2, pp. 336–377.

Lee, Sang Yoon (Tim) and Ananth Seshadri (2019). «On the Intergenerational Transmission of Economic Status». In: *Journal of Political Economy* 127.2, pp. 855–921.

Lenoël, Audrey and Anda David (2019). «Leaving Work Behind? The Impact of Emigration on Female Labor Force Participation in Morocco». In: *International Migration Review* 53.1, pp. 122–153.

Leukhina, Oksana M. and Stephen J. Turnovsky (2016). «Population Size Effects in the Structural Development of England». In: *American Economic Journal: Macroeconomics* 8.3, pp. 195–229.

Leuven, Edwin and Barbara Sianesi (2003). *PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing*. Statistical Software Components, Boston College Department of Economics.

Lew, Byron and Bruce Cater (2018). «Farm mechanization on an otherwise ‘featureless’ plain: tractors on the Northern Great Plains and immigration policy of the 1920s». In: *Cliometrica* 12.2, pp. 181–218.

Lewis, Ethan (2011). «Immigration, Skill Mix, and Capital Skill Complementarity». In: *The Quarterly Journal of Economics* 126.2, pp. 1029–1069.

— (2013). «Immigration and Production Technology». In: *Annual Review of Economics* 5.1, pp. 165–191.

Lewis, W. Arthur (1954). «Economic Development with Unlimited Supplies of Labour». In: *The Manchester School* 22.2, pp. 139–191.

Lingwall, Jeff (2010). «Compulsory Schooling, the Family, and the 'Foreign Element' in the United States, 1880–1900». Unpublished manuscript.
Lochner, Lance and Alexander Monge-Naranjo (2012). «Credit Constraints in Education». In: *Annual Review of Economics* 4.1, pp. 225–256.
Løken, Katrine V. (2010). «Family income and children's education: Using the Norwegian oil boom as a natural experiment». In: *Labour Economics* 17.1, pp. 118–129.
Løken, Katrine V., Magne Mogstad, and Matthew Wiswall (2012). «What Linear Estimators Miss: The Effects of Family Income on Child Outcomes». In: *American Economic Journal: Applied Economics* 4.2, pp. 1–35.
Long, Jason and Joseph Ferrie (2007). «The Path to Convergence: Intergenerational Occupational Mobility in Britain and the US in Three Eras*». In: *The Economic Journal* 117.519, pp. C61–C71.
— (2013). «Intergenerational Occupational Mobility in Great Britain and the United States since 1850». In: *American Economic Review* 103.4, pp. 1109–37.
López, Ricardo A. and Jens Südekum (2009). «Vertical Industry Relations, Spillovers, And Productivity: Evidence From Chilean Plants». In: *Journal of Regional Science* 49.4, pp. 721–747.
Lucas, Adrienne M. and Isaac M. Mbiti (2012). «Access, Sorting, and Achievement: The Short-Run Effects of Free Primary Education in Kenya». In: *American Economic Journal: Applied Economics* 4.4, pp. 226–253.
Mailáth, József (1900). «Az amerikai kérdésröl - gróf Mailáth beszéde ». In: *Zemplén* 31.19, pp. 1–3.
Malamud, Ofer and Abigail Wozniak (2012). «The Impact of College on Migration: Evidence from the Vietnam Generation». In: *Journal of Human Resources* 47.4, pp. 913–950.
Maloney, William F and Felipe Valencia Caicedo (2020). *Engineering Growth*. CEPR Discussion Papers 15144. C.E.P.R. Discussion Papers.
Manoli, Day and Nicholas Turner (2018). «Cash-on-Hand and College Enrollment: Evidence from Population Tax Data and the Earned Income Tax Credit». In: *American Economic Journal: Economic Policy* 10.2, pp. 242–271.
Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles (2021). *IPUMS National Historical Geographic Information System: Version 16.0 [dataset].*
Margo, Robert A. (2000). *Wages and Labor Markets in the United States, 1820-1860*. National Bureau of Economic Research Books 9780226505077. University of Chicago Press. ISBN: ARRAY(0x4f871030).
Marshall, George (1890). *Principles of Economics*. MacMillan & Co.
Martínez-Galarraga, Julio (2012). «The determinants of industrial location in Spain, 1856–1929». In: *Explorations in Economic History* 49.2, pp. 255–275.

Martínez-Galarraga, Julio, Elisenda Paluzie, Jordi Pons, and Daniel A. Tirado-Fabregat (2008). «Agglomeration and labour productivity in Spain over the long term». In: *Cliometrica, Journal of Historical Economics and Econometric History* 2.3, pp. 195–212.

Matsuyama, Kiminori (1991). «Increasing Returns, Industrialization, and Indeterminacy of Equilibrium». In: *The Quarterly Journal of Economics* 106.2, pp. 617–650.

— (1992). «Agricultural productivity, comparative advantage, and economic growth». In: *Journal of Economic Theory* 58.2, pp. 317–334.

Mayr, Karin and Giovanni Peri (2009). «Brain Drain and Brain Return: Theory and Application to Eastern-Western Europe». In: *The B.E. Journal of Economic Analysis & Policy* 9.1, pp. 1–52.

McKenzie, David and Hillel Rapoport (2010). «Self-Selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks». In: *The Review of Economics and Statistics* 92.4, pp. 811–821.

Meisenzahl, Ralf R. and Joel Mokyr (2011). «The Rate and Direction of Invention in the British Industrial Revolution: Incentives and Institutions». In: *The Rate and Direction of Inventive Activity Revisited*. NBER Chapters. National Bureau of Economic Research, Inc, pp. 443–479.

Mendola, Mariapia and Calogero Carletto (2012). «Migration and gender differences in the home labour market: Evidence from Albania». In: *Labour Economics* 19.6, pp. 870–880.

MGYOSZ (1907). *A kivándorlás. A Magyar Gyáriparosok Országos Szövetsége által tartott országos ankét tárgyalásai*. Athenaeum.

Mian, Atif and Amir Sufi (2014). «What Explains the 2007–2009 Drop in Employment?» In: *Econometrica* 82, pp. 2197–2223.

Mishra, Prachi (2014). «Emigration and wages in source countries: a survey of the empirical literature». In: *International Handbook on Migration and Economic Development*. Ed. by Robert E.B. Lucas. Chapters. Edward Elgar Publishing. Chap. 9, pp. 241–266.

Mogstad, Magne and Gaute Torsvik (2021). *Family Background, Neighborhoods and Intergenerational Mobility*. NBER Working Papers 28874. National Bureau of Economic Research, Inc.

Mokyr, Joel (1999a). «The Second Industrial Revolution, 1870-1914». In: *Storia dell'economia Mondiale*. Ed. by Valerio Castronovo. Laterza Publishing, pp. 219–245.

— (1999b). «The Second Industrial Revolution, 1870-1914». In: *Storia dell'economia Mondiale*. Ed. by Valerio Castronovo. Laterza Publishing, pp. 219–245.

Mokyr, Joel and Hans-Joachim Voth (2010). «Understanding growth in Europe, 1700-1870: theory and evidence». In: *The Cambridge Economic History of*

Modern Europe. Ed. by Stephen Broadberry and Kevin H. O'Rourke. Vol. 1. Cambridge University Press, pp. 7–42.
Molina, Teresa and Ivan Rivadeneyra (2021). «The schooling and labor market effects of eliminating university tuition in Ecuador». In: *Journal of Public Economics* 196.C.
Mörk, Eva, Anna Sjögren, and Helena Svaleryd (2020). «Consequences of parental job loss on the family environment and on human capital formation-Evidence from workplace closures». In: *Labour Economics* 67.C.
Munshi, Kaivan (2020). «Social Networks and Migration». In: *Annual Review of Economics* 12.1, pp. 503–524.
Munshi, Kaivan and Mark Rosenzweig (2016). «Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap». In: *American Economic Review* 106.1, pp. 46–98.
Murphy, Kevin M, Andrei Shleifer, and Robert W Vishny (1989). «Industrialization and the Big Push». In: *Journal of Political Economy* 97.5, pp. 1003–1026.
Nagy, Dávid Krisztián (forthcoming). «Trade and Urbanization: Evidence from Hungary ». In: *American Economic Journal: Microeconomics*.
Nagy, Valéria (1983). «A kivándorlást gátló tényezők Hajdu-Bihar megyében a századfordulón». In: *Történelmi szemle* XXXVI.1, pp. 176–187.
Neilson, Christopher A. and Seth D. Zimmerman (2014). «The effect of school construction on test scores, school enrollment, and home prices». In: *Journal of Public Economics* 120.C, pp. 18–31.
Neményi, Bertalan (1911). *A magyar nép állapota és az amerikai kivándorlás*. Athenaeum.
Ngai, L. Rachel and Christopher A. Pissarides (2007). «Structural Change in a Multisector Model of Growth». In: *American Economic Review* 97.1, pp. 429–443.
Nikolić, Stefan (2018). «Determinants of industrial location: Kingdom of Yugoslavia in the interwar period». In: *European Review of Economic History* 22.1, pp. 101–133.
O'Brien, Anthony Patrick (1988). «Factory size, economies of scale, and the great merger wave of 1898-1902». In: *The Journal of Economic History* 48.3, pp. 639–649.
Olivetti, Claudia and M. Daniele Paserman (2015). «In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940». In: *American Economic Review* 105.8, pp. 2695–2724.
Oreopoulos, Philip, Marianne Page, and Ann Huff Stevens (2008). «The Intergenerational Effects of Worker Displacement». In: *Journal of Labor Economics* 26.3, pp. 455–483.

Oster, Emily (2019). «Unobservable Selection and Coefficient Stability: Theory and Evidence». In: *Journal of Business & Economic Statistics* 37.2, pp. 187–204.

Pálvölgyi, Balázs (2010). «A magyar kivándorlás politikai kezdetei (1881–1903) – Kivándorlási törvények és az amerikai kivándorlás». In: *Jogtörténeti Szemle* 12.27-34.

— (2018). «The Changes of the Hungarian Migration Policy in the First 25 Years of the 20th Century». In: *Pro Publico Bono - Magyar Közigazgatás* 1.204-222.

Papageorgiou, Theodore (2014). «Learning Your Comparative Advantages». In: *Review of Economic Studies* 81.3, pp. 1263–1295.

Parman, John (2011). «American Mobility and the Expansion of Public Education». In: *The Journal of Economic History* 71.1, pp. 105–132.

Peters, Michael (2021). «Market Size and Spatial Growth - Evidence from Germany's Post-War Population Expulsions». Unpublished manuscript.

Poznan, Kristina E. (2017). «Return Migration to Austria-Hungary from the United States in Homeland Economic and Ethnic Politics and International Diplomacy». In: *The Hungarian Economic Review* 6.3.

Preston, Samuel H. and Michael R. Haines (1991). *Fatal Years: Child Mortality in Late Nineteenth-Century America*. NBER Books pres91-1. National Bureau of Economic Research, Inc.

Puskás, Julianna (1982). *Kivándorló magyarok az Egyesült Államokban 1880–1940*. Akadémiai Kiadó.

— (1983). «Hungarian Migration Patterns - New Research in Hungary». In: *Acta Historia Academiae Scientiarium Hungaricae* 29.2/4, pp. 265–272.

Rege, Mari, Kjetil Telle, and Mark Votruba (2011). «Parental Job Loss and Children's School Performance». In: *Review of Economic Studies* 78.4, pp. 1462–1489.

Roodman, D., J. MacKinnon, M. Nielsen, and M. Webb (2018). «Fast and wild: bootstrap inference in Stata using boottest». In: *Queen's Economics Department Working Paper* 1406.

Rosenbaum, Paul R. and Donald B. Rubin (1983). «The central role of the propensity score in observational studies for causal effects». In: *Biometrika* 70.1, pp. 41–55.

Rosenberg, Nathan and Manuel Trajtenberg (2004). «A General-Purpose Technology at Work: The Corliss Steam Engine in the Late-Nineteenth-Century United States». In: *The Journal of Economic History* 64.1, pp. 61–99.

Rosenbloom, Joshua L. (1990). «One Market or Many? Labor Market Integration in the Late Nineteenth-Century United States». In: *The Journal of Economic History* 50.1, pp. 85–107.

Rosenbloom, Joshua L. (1996). «Was There a National Labor Market at the End of the Nineteenth Century? New Evidence on Earnings in Manufacturing». In: *The Journal of Economic History* 56.3, pp. 626–656.

— (1998). «The Extent of the Labor Market in the United States, 1870-1914». In: *Social Science History* 22.3, pp. 287–318.

— (2002). *Looking for Work, Searching for Workers: American Labor Markets during Industrialization*. Cambridge University Press.

Rosenbloom, Joshua L. and William A. Sundstrom (2003). *The Decline and Rise of Interstate Migration in the United States: Evidence from the IPUMS, 1850-1990*. NBER Working Papers 9857. National Bureau of Economic Research, Inc.

Rosenthal, Stuart S. and William C. Strange (2004). «Evidence on the nature and sources of agglomeration economies». In: *Handbook of Regional and Urban Economics*. Ed. by J. V. Henderson and J. F. Thisse. Vol. 4. Handbook of Regional and Urban Economics. Elsevier. Chap. 49, pp. 2119–2171.

Rosés, Joan R. (2003). «Why Isn't the Whole of Spain Industrialized? New Economic Geography and Early Industrialization, 1797-1910». In: *The Journal of Economic History* 63.4, pp. 995–1022.

Ruggles, Steven, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler, and Matthew Sobek (2021). *IPUMS USA: Version 11.0 [dataset]*.

Saavedra, Martin and Tate Twinam (2020). «A machine learning approach to improving occupational income scores». In: *Explorations in Economic History* 75.C.

Salhofer, Klaus (2001). «Elasticities of Substitution and Factor Supply Elasticities in European Agriculture: A Review of Past Studies». In: *Market Effects of Crop Support Measures*. Paris: OECD, pp. 89–119.

Sanders, Carl and Christopher Taber (2012). «Life-Cycle Wage Growth and Heterogeneous Human Capital». In: *Annual Review of Economics* 4.1, pp. 399–425.

Santangelo, Gabriella (2019). *Firms and Farms: The Local Effects of Farm Income on Firms' Demand*. Cambridge Working Papers in Economics. Faculty of Economics, University of Cambridge.

Sarvimäki, Matti (2011). *Agglomeration in the periphery*. LSE Research Online Documents on Economics 57857. London School of Economics and Political Science, LSE Library.

Schulze, Max-Stephan (2007a). «Origins of catch-up failure: Comparative productivity growth in the Habsburg Empire, 1870–1910». In: *European Review of Economic History* 11.2, pp. 189–218.

— (2007b). *Regional income dispersion and market potential in the late nineteenth century Hapsburg Empire*. Economic History Working Papers 22311. London School of Economics and Political Science, Department of Economic History.

Schulze, Max-Stephan and Nikolaus Wolf (2012). «Economic nationalism and economic integration: the Austro-Hungarian Empire in the late nineteenth century». In: *Economic History Review* 65.2, pp. 652–673.

Sequeira, Sandra, Nathan Nunn, and Nancy Qian (2020). «Immigrants and the Making of America». In: *Review of Economic Studies* 87.1, pp. 382–419.

Sheridan, Frank Joseph (1907). «Italian, Slavic and Hungarian unskilled immigrants laborers in the United States». In: *Bulletin of the Bureau of Labour* XV.72, pp. 403–486.

Shrestha, Slesh A. (2017). «No Man Left Behind: Effects of Emigration Prospects on Educational and Labour Outcomes of Non-migrants». In: *Economic Journal* 127.600, pp. 495–521.

Sobek, Matthew (1996). «Work, Status, and Income: Men in the American Occupational Structure since the Late Nineteenth Century». In: *Social Science History* 20.2, pp. 169–207.

Sokoloff, Kenneth L. and B. Zorina Khan (1990). «The Democratization of Invention During Early Industrialization: Evidence from the United States, 1790-1846». In: *The Journal of Economic History* 50.2, pp. 363–378.

Solis, Alex (2017). «Credit Access and College Enrollment». In: *Journal of Political Economy* 125.2, pp. 562–622.

Solon, Gary (2004). «A model of intergenerational mobility variation over time and place». In: *Generational Income Mobility in North America and Europe*. Ed. by Miles Corak. Cambridge University Press, pp. 38–47.

Song, Xi, Catherine G. Massey, Karen A. Rolf, Joseph P. Ferrie, Jonathan L. Rothbaum, and Yu Xie (2020). «Long-term decline in intergenerational mobility in the United States since the 1850s». In: *Proceedings of the National Academy of Sciences* 117.1, pp. 251–258.

Spitzer, Yannay and Ariell Zimran (2018). «Migrant self-selection: Anthropometric evidence from the mass migration of Italians to the United States, 1907–1925». In: *Journal of Development Economics* 134.C, pp. 226–247.

— (2019). «Like an Ink Blot on Paper: Testing the Diffusion Hypothesis of Mass Migration, Italy 1876–1920». Unpublished manuscript.

Squicciarini, Mara P. and Nico Voigtländer (2015). «Human Capital and Industrialization: Evidence from the Age of Enlightenment». In: *The Quarterly Journal of Economics* 130.4, pp. 1825–1883.

Stecker, Margaret (1937). *Intercity differences in costs of living in March, 1935, 59 cities*. Washington: U.S. Govt. Print. Office.

Stokey, Nancy L (1996). «Free Trade, Factor Returns, and Factor Accumulation». In: *Journal of Economic Growth* 1.4, pp. 421–447.

Sundstrom, William A. and Joshua L. Rosenbloom (1993). «Occupational Differences in the Dispersion of Wages and Working Hours: Labor Market Integration

in the United States, 1890-1903». In: *Explorations in Economic History* 30.4, pp. 379–408.
Swiecki, Tomasz (2017). «Determinants of Structural Change». In: *Review of Economic Dynamics* 24, pp. 95–131.
Tabellini, Marco (2020). «Gifts of the Immigrants, Woes of the Natives: Lessons from the Age of Mass Migration». In: *Review of Economic Studies* 87.1, pp. 454–486.
Tonelli, Sándor (1908). «Utazás a magyar kivándorlókkal Amerikába». In: *Közgazdasági Szemle* 40, pp. 427–444.
Traiberman, Sharon (2019). «Occupations and Import Competition: Evidence from Denmark». In: *American Economic Review* 109.12, pp. 4260–4301.
Tyack, David (1974). *The one best system : a history of American urban education*. Cambridge, Mass. : Harvard University Press.
U.S. Department of Labor (1899). *Thirteenth Annual Report of the Commissioner of Labor*. Vol. 2. Washington: U.S. G.P.O.
— (1900). *Fifteenth Annual Report of the Commissioner of Labor*. Washington: U.S. G.P.O.
Wantchekon, Leonard, Marko Klašnja, and Natalija Novta (2015). «Education and Human Capital Externalities: Evidence from Colonial Benin». In: *The Quarterly Journal of Economics* 130.2, pp. 703–757.
Ward, Zachary (2019). *Intergenerational Mobility in American History: Accounting for Race and Measurement Error*. CEH Discussion Papers 10. Centre for Economic History, Research School of Economics, Australian National University.
— (2020). «The Not-So-Hot Melting Pot: The Persistence of Outcomes for Descendants of the Age of Mass Migration». In: *American Economic Journal: Applied Economics* 12.4, pp. 73–102.
— (forthcoming). «Internal Migration, Education and Upward Rank Mobility:Evidence from American History». In: *The Journal of Human Resources*.
Weil, David N. and Joshua Wilde (2009). «How Relevant Is Malthus for Economic Development Today?» In: *American Economic Review* 99.2, pp. 255–260.
Williamson, Jeffrey G. and Peter H. Lindert (1980). «Long-Term Trends in American Wealth Inequality». In: *Modeling the Distribution and Intergenerational Transmission of Wealth*. NBER Chapters. National Bureau of Economic Research, Inc, pp. 9–94.
Wozniak, Abigail (2010). «Are College Graduates More Responsive to Distant Labor Market Opportunities?» In: *Journal of Human Resources* 45.4, pp. 944–970.
Wright, Nicholas A. (2021). «Need-based financing policies, college decision-making, and labor market behavior: Evidence from Jamaica». In: *Journal of Development Economics* 150.C.

Wyman, Mark (1993). *Round-trip to America: The Immigrants Return to Europe, 1880-1930*. Cornell University Press.
Xie, Bin (2017). *The Effects of Immigration Quotas on Wages, the Great Black Migration, and Industrial Development*. IZA Discussion Papers 11214. Institute of Labor Economics (IZA).
Yang, Dean (2011). «Migrant Remittances». In: *Journal of Economic Perspectives* 25.3, pp. 129–152.
Zahra, Tara (2017). *The Great Departure: Mass Migration from Eastern Europe and the Making of the Free World*. W. W. Norton Company.
Zarnowitz, Victor (1996). *Business Cycles: Theory, History, Indicators, and Forecasting*. University of Chicago Press.