Immigrants and the Making of America: A Comment*

Joop Adema†

May 17, 2026

Abstract

Sequeira, Nunn and Qian (2020) study the long-run impact of immigration during the Age of Mass Migration on U.S. county-level economic outcomes in 2000 and report large positive effects. They instrument historical immigration with the interaction between county railroad connectivity and European migration to the U.S. averaged over seven decades (1860–1920). Because this instrument predicts larger inflows in earlier-connected counties, the authors state that they control for this by controlling for *log(years since railroad connection in 2000)*. However, to accommodate the 4% of counties that never received access, they instead control for *log(years since railroad connection in 2000 + 1)*, which mainly captures differences between ever- and never-connected counties, but not between earlier and later connected counties. In this Comment, I show that the main results disappear when the timing of railroad connection is parametrically controlled for, suggesting the original findings reflect the better long-run performance of early-connected counties rather than immigration. Turning to the recent literature on shift–share instruments, I further demonstrate that asymptotic inference severely overstates precision due to the small number of shifts (seven). The instrument thus provides insufficient identifying variation to credibly estimate the long-run effects of migration.

1 Introduction

Identifying the long-run effects of historical shocks is challenging. A possible research design isolates historical shocks, aggregates them and examines cross-sectional differences in outcomes in a later time period. A recent and prominent example is Sequeira, Nunn and Qian (2020) (hereafter: SNQ). They study how migration to US counties during the Age of Mass Migration (1860-1920) shapes modern-day (2000) economic outcomes. SNQ instrument for the county-level migration share using the interaction between a binary indicator for railroad connectivity and the total immigration share to the US, averaged across all decades the county existed between 1860 and 1920. They report large effects: a five-percentage-point increase in the average migrant share raises per capita income in 2000 by 13%. They also find that residents of counties with higher historical migration attain more years of schooling.

In this Comment, I show that these findings are the consequence of inadequately controlling for the timing of connection to the railroad network. SNQ's instrument predicts more

*I thank the editor, Katrine V. Løken, two anonymous referees for insightful comments and the original authors for their response and the availability of replication data and code. I downloaded the data on May 18, 2025 from `https://nathannunn.arts.ubc.ca/data/`. All code to replicate my reanalysis can be accessed here: `https://jopieadema.github.io/replication/replication_SNQ.zip`

†Post-doc at the University of Innsbruck, email: joop.adema@uibk.ac.at

migration to counties connected earlier, which are plausibly more prosperous in the long run for reasons unrelated to migration, threatening the exclusion restriction. SNQ acknowledge this and control parametrically for timing of railroad connection by including a covariate for the log number of years since railroad connection in 2000 (*log(years of rail+1)*).[1] However, inspection of SNQ's replication package reveals that their implementation differs. As 127 (4%) counties in the sample never received railways, the authors assign a value of 0 for those counties and use the log(x+1) transformation instead. This control largely captures differences between ever- and never-connected counties rather than between earlier and later connected counties. If counties connected earlier to railroads are systematically more prosperous today, the exclusion restriction may be violated, leading to an upward bias in the estimate. After controlling for time since railroad connection, I find that the estimates of the impact on economic and educational outcomes turn negative and insignificant.

Do these corrected estimates credibly identify the long-run causal effect of migration? To address this question, I draw on the recent literature on shift–share IVs (SSIV). I show that the instrument can be written as a SSIV, where the shifts are seven decadal migration shocks and the shares are binary indicators for railroad connection in a given decade divided by the total number of decades a county existed between 1860-1920. A linear shift–share instrument is systematically correlated to the sum of the shares, which is here the ratio between (i) the number of decades a county had railroad connection and (ii) the number of decades a county existed during the Age of Mass Migration. This also reveals an additional correlation: the instrument is systematically higher for younger counties. Identification can be achieved by controlling for the sum of shares.

Controlling for the sum of shares leaves a moderately strong first stage, but renders the main estimate considerably smaller and insignificant. However, the residualized instrument still correlates in a non-linear fashion to the timing of railroad connection and county age due to the limited number of shifts. Including fixed effects for the decade of railroad connection turns the estimated effect on income per capita in 2000 negative, while including fixed effects for the decade of county establishment renders the first stage irrelevant. Hence, the instrument lacks sufficient identifying variation unless one is willing to assume that the timing of county establishment is unrelated to long-run outcomes.

The limited number of shifts also provides a serious threat to inference: performing a randomization inference procedure by permuting the decadal shocks shows that SNQ's spatial standard errors severely overestimate precision. In the original specification, randomly permuting the seven decadal migration shocks and recalculating the instrument yields a positive and statistically significant effect in the reduced form regression in 53% of cases ($\alpha = 0.05$). In the specification that controls for the sum of shares, the reduced form results are not biased towards positive estimates, but asymptotic inference rejects the null in 61% of cases ($\alpha = 0.05$). Consequently, the 90% confidence interval of the effect of a 1 percentage point higher migration share during the Age of Mass Migration on per capita incomes in 2000 is uninformatively large: it ranges from a reduction of 12% to an increase of 8%. I conclude that the approach of SNQ cannot be used to answer the research question at hand, independent of exact identification assumptions.

The remainder of the paper proceeds as follows. Section 2 discusses SNQ's empirical approach, Section 3 examines the control for time since railroad connection and Section 4

[1] SNQ write (p. 384): *"One concern with our identification strategy is that the interaction of connection to the railway network and aggregate immigrant inflows might be correlated with how early a county became connected to the railway. To address this, I always control for a measure of how early the county became connected to the railway."*

discusses identification and inference through the lens of the shift–share literature. Section 5 concludes.

2 Empirical approach

SNQ aim to study the causal effect of immigration between 1860 and 1920 on county-level outcomes in the year 2000. As county-level immigration is likely endogenously determined, SNQ use an instrumental variables approach that uses the interaction between whether a county i has railroads and total US-wide immigrant inflows in the decade before census year t, aggregated over the seven decades between 1860 and 1920. The rationale behind the instrument is that counties that were connected prior to a large influx of immigrants are expected to receive more migrants. SNQ estimate the following two-stage least squares (2SLS) procedure:

$$\text{Avg Immigrant Share}_{i,s} = \zeta_s + \mu \widehat{\text{Avg Immigrant Share}}_{i,s} + \omega \text{RR Duration}_{i,s} + X_{i,s}\Omega + \epsilon_{i,s} \tag{1}$$

$$Y_{i,s} = \xi_s + \psi \text{Avg Immigrant Share}_{i,s} + \pi \text{RR Duration}_{i,s} + X_{i,s}\Pi + \nu_{i,s} \tag{2}$$

where i indexes a county in state s. Avg Immigrant $\text{Share}_{i,s}$ is the average migration share during 1860-1920. The instrument for the immigrant share is given by:

$$\widehat{\text{Avg Immigrant Share}}_{i,s} = \frac{1}{T_i} \sum_{t=T_i^0+10}^{1920} \text{Immigrant Flow}_{t-10} \times I_{i,t-10}^{\text{RR Access}} \tag{3}$$

The notation here slightly differs from that in SNQ, to make it explicit that the period is a decade and that the panel is unbalanced.[2] T_i^0 denotes the first decade county i appears in the panel used to construct the instrument, which I refer to as the decade of establishment. $T_i = (1920 - T_i^0)/10$ denotes the number of decades the county existed during the Age of Mass Migration. Immigrant Flow_{t-10} is the ratio between total decadal migration to the US between $t-10$ to $t-1$ and the total population of the US in census year $t-10$. $I_{i,t-10}^{\text{RR Access}}$ is a binary indicator for whether a county had railroad access at $t-10$. The instrument can be rewritten as a shift–share IV:

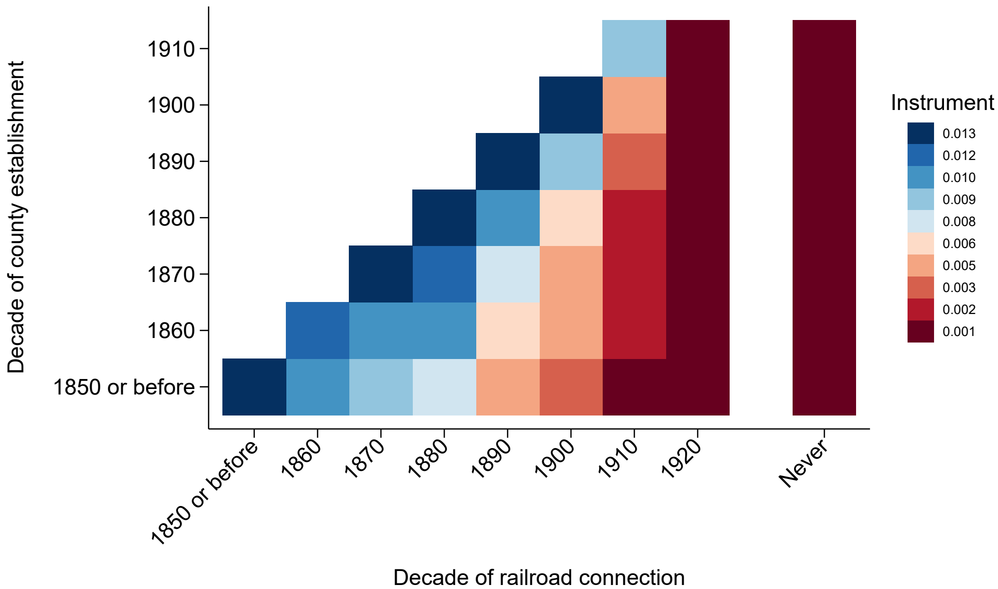
$$\widehat{\text{Avg Immigrant Share}}_{i,s} = \sum_{t=1860}^{1920} \text{Immigrant Flow}_{t-10} \times RR_{i,t-10} \tag{4}$$

where Immigrant Flow_{t-10} are the seven decadal shifts. The shares are given by:

$$RR_{i,t-10} = \begin{cases} \dfrac{I_{i,t-10}^{\text{RR Access}}}{T_i}, & \text{if } t-10 \geq T_i^0, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

[2]SNQ write (p. 391): *Since some counties were still in the process of being formed during this period, our panel is unbalanced with counties entering over time. In 1860, there are 1,532 counties in our sample, there are 1,922 counties in 1870; 2,137 in 1880; 2,416 in 1890; 2,692 in 1900; 2,752 in 1910; and 2,935 in 1920.* Counties established during the 1920s or later are not part of the sample.

Figure 1: Instrument value as a function of establishment and railroad cohorts



Notes: Instrument value by decade of establishment and decade of railroad connection. N = 2,935.

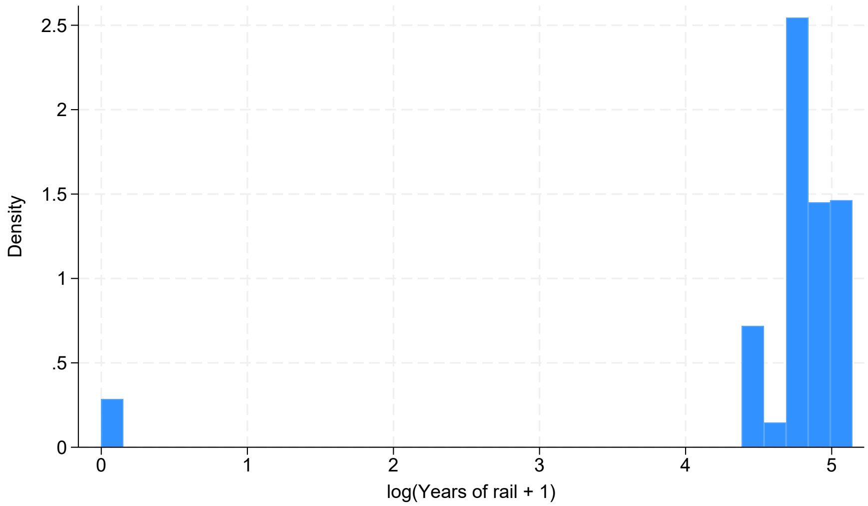
The share equals zero prior to railroad connection and takes the value $1/T_i$ from the first decade after railroad connection onward. The value of the instrument is determined at the level of decade of railroad connection by decade of county establishment, shown in Figure 1. For two counties established in the same decade, earlier railroad connection implies a higher instrument value because $RR_{i,t-10} = 1$ for more periods.[3] Due to pre-existing differences or better future potential, early connection is plausibly associated with better long-run economic outcomes. This motivates SNQ to control parametrically for the timing of railroad connection, RR $\text{Duration}_{i,s}$, which we discuss in Section 3.

Similarly, for two counties connected in the same decade but established at different times, the older county has a lower instrument value because T_i is larger. The instrument is therefore also mechanically correlated with county age, which may also be related to long-run economic outcomes—for example, if older counties were located in areas more favorably positioned to benefit from industrialization. SNQ do not acknowledge this mechanical correlation or discuss its implications for identification.

SNQ introduce several additional controls important for identification. They include state fixed effects (ζ_s and ξ_s) and flexible controls for latitude and longitude to capture geography and historical factors. In addition, they include two controls that mirror the structure of the instrument in Equation 4: the shares are identical, but the shifts are replaced by (i) the log of decadal US-wide industrialization and (ii) decadal US-wide GDP growth. These variables are included to capture differences in long-run economic performance of counties who had a railroad connection during periods of greater industrialization or strong economic growth.

[3] In SNQ's data, railroad connection is an absorbing state: once connected, a county remains connected.

Figure 2: Histogram of *log(years of rail+1)*



Notes: Histogram of the *log(years of rail+1)* control. N = 2,935.

3 Controlling for time since railroad connection

3.1 SNQ's implementation

SNQ underscore the importance of controlling for the timing of connection to the railroad network (p. 392): "*$RR\_Duration_{is}$ is the log number of years, as of 2000, that a county has been connected to the railroad network. The variable is included to address the possibility that our instrument may be correlated with early connection to the railroad network, which could have an independent long-run effect on our outcomes of interest.*"

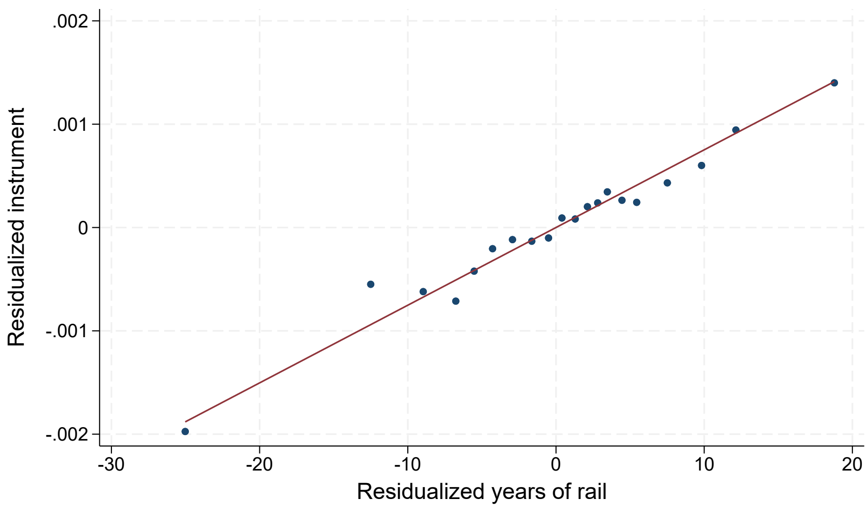
However, the authors do not perform the log transformation as stated in the quote, but assign counties that did not receive a railroad connection by 1930 a value of 0 and use the *log(x+1)* transformation instead. This ad hoc choice was likely made to accommodate never connected counties; in doing so, SNQ treat whether a county never received railroads as qualitatively similar to when a county received railroads. This concerns 127 out of 2,935 counties (4%).[4]

Figure 2 shows the distribution of the *log(years of rail+1)* control. All 127 counties that never received a railroad connection are at 0 and all other counties are in a narrow range between 4.4 and 5.1. This shows that it predominantly captures differences between counties that never had a railroad connection and those that did (the extensive margin), rather than controlling for *when* a county got connected to the railroads, as the authors intended to do (the intensive margin). The correlation coefficient between *log(years of rail+1)* and a binary indicator for the 127 never-connected counties is $\rho = -0.985$.

The shift–share-like control functions for industrialization and GDP growth could, in principle, absorb effects related to the timing of railroad connection. However, the decadal shocks underlying these controls are only weakly correlated with the decadal migration shocks (see Online Appendix Figure A1) and therefore may only partially account for the

[4]SNQ only mention the presence of never-connected counties when introducing two additional instruments in Section 7.2, but do not discuss it when introducing the control.

Figure 3: Binned scatterplot of the residualized instrument and years since railroad connection



Notes: Binned scatterplot ($N_{bin} = 20$) with trendline of the residualized instrument and the residualized number of years since railroad connection. The control variables used for residualization are the same as in column 1 of Table 4 of SNQ. N = 2,935.

instrument's dependence on railroad timing. Online Appendix Figure A2a plots the average instrument and control values by decade of railroad connection. The average instrument is lower for counties connected in later decades, whereas the two control functions exhibit a rather flat trend until 1910.[5] As a result, these controls mainly distinguish the 1920 cohort from earlier connected cohorts, but do not capture gradual differences across counties connected between 1830 and 1910.

To examine the potential consequences of not explicitly controlling for the intensive margin of railroad connection, I residualize (i) the instrument and (ii) the number of years since connection in 2000 using the full set of controls in Equation 2, including the *log(years of rail+1)* control and shift–share-like control functions. Figure 3 presents the residual correlation in a binned scatterplot: the residualized instrument is strongly correlated with residualized years since connection. If early- and late-connected counties followed different long-run economic trajectories, the exclusion restriction is likely violated.

Figures A3a and A4a plot the standardized residualized instrument by decade of railroad connection and decade of county establishment for different specifications; the legends report the corresponding residual correlations. In line with Figure 3, the residualized instrument is higher in earlier-connected counties. Figures A3b and A4b present analogous plots for residualized per capita income levels in 2000. Both the residualized instrument and residualized income per capita are lower for later connected and younger cohorts, consistent with an upward bias in the causal effect of migration on incomes in 2000. Including *log(years of rail+1)* and an indicator for never-connected counties controls both the extensive margin and the intensive margin (with a specific functional form) of railroad connection; I implement this specification in Section 3.2.

[5]Instrument and control function values are 0 for the 1920 and never-connected cohorts as Equation 3 uses the first lag of railroad access.

Figures A3 and A4 also show that substantial, non-linear variation in the instrument across cohorts remains even after parametrically controlling for both the extensive and intensive margin of railroad connection. Residual variation in the instrument is particularly pronounced for the 1910 and 1920 railroad cohorts and the 1900 and 1910 establishment cohorts. A causal interpretation therefore requires assuming that unobserved determinants of long-run outcomes are unrelated to these cohort-specific patterns. Section 4.2 explores alternative parametric and more flexible controls for connection timing and county age.

3.2 Re-analysis

Table 1 shows how the first stage, reduced form, and 2SLS estimates change for different sets of controls. I focus on SNQ's main outcome—log income per capita in 2000 (column 1 of SNQ Table 4)—and discuss other outcomes below. Column 1 replicates SNQ's baseline specification, which includes all baseline controls including the *log(years of rail+1)* control. Column 2 instead replaces this term with an indicator for counties that never received railroad access. Reduced form and 2SLS results are nearly unchanged, which is unsurprising since in isolation both controls mainly capture the extensive margin of railroad connection. The coefficient on the control indicates little difference in long-run outcomes between counties that ever received a railroad and those that never did.

Table 1: Re-analysis of SNQ Table 4, Column (1)

Sample:	Full sample				Ever connected by rail	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: First stage	Average immigrant share, 1860–1920					
Instrument	4.559^{***}	4.512^{***}	3.929^{**}	3.953^{**}	3.788^{**}	4.524^{***}
	(1.311)	(1.290)	(1.810)	(1.753)	(1.854)	(1.318)
Panel B: Reduced Form	Log income per capita, 2000					
Instrument	11.942^{***}	12.254^{***}	-7.767^{*}	-8.042^{*}	-8.775^{**}	12.641^{***}
	(3.629)	(3.626)	(4.474)	(4.584)	(4.357)	(3.596)
Panel C: 2SLS	Log income per capita, 2000					
Average Immigrant	2.619^{***}	2.716^{***}	-1.977	-2.034^{*}	-2.316^{*}	2.794^{***}
Share 1860-1920	(1.022)	(1.046)	(1.452)	(1.435)	(1.639)	(1.063)
RI p-value	{0.791}	{0.767}	{0.436}	{0.420}	{0.404}	{0.797}
90% Anderson-Rubin CI	[1.26, 5.28]	[1.35, 5.43]	[-9.09, -0.12]	[-8.28, -0.15]	[-13.02, -0.39]	[1.41, 5.64]
log(Years of rail + 1)	0.005		0.348^{***}		0.386^{***}	
	(0.006)		(0.081)		(0.079)	
Never connected		-0.013	1.571^{***}	0.266^{***}		
		(0.030)	(0.369)	(0.073)		
Years of rail				0.003^{***}		
				(0.001)		
Observations	2935	2935	2935	2935	2808	2808
First stage F	12.09	12.23	4.71	5.09	4.17	11.78
R^2 of IV on controls	0.931	0.930	0.963	0.963	0.951	0.908

Notes: First stage, reduced form and 2SLS estimates based on column 1 of Table 4 of SNQ. Column 1 presents the original estimates and columns 2-4 introduce different sets of control variables (as shown in Panel C). Column 5 and 6 exclude the 127 never-connected counties; column 5 includes all baseline controls, whereas column 6 removes the control for *log(years of rail+1)*. Coefficients on additional controls shown in Panel C are omitted for brevity in Panel A and B. Conley standard errors using a five-degree window are shown in parentheses. Anderson-Rubin confidence intervals (square brackets) and Randomization Inference p-values (curly brackets) are reported for the 2SLS coefficient of interest. Kleibergen-Paap F-statistics based on Conley standard errors are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Column 3 includes both controls simultaneously. Conditional on the indicator for ever receiving a railroad, *log(years of rail+1)* now captures the timing of railroad connection with a particular functional form. As the first stage becomes substantially weaker, I report 90% Anderson–Rubin confidence intervals (AR CIs), which remain valid even in the presence of weak instruments.[6] The reduced form and 2SLS estimates reverse sign and become statistically insignificant, but the 90% confidence interval does not include 0. The coefficient on the *log(years of rail+1)* control also explains SNQ's main result: counties with rail access 10% longer have about 3.3% higher per capita income in 2000, indicating that the timing of railroad connection—not migration—drives the positive estimate in column 1. Column 4 replaces *log(years of rail+1)* with a linear timing control, which yields similar conclusions: the instrumented coefficient is negative and marginally significant, while receiving rail access 10 years earlier is associated with roughly 3% higher income.

Column 5 reports results from the baseline specification in column 1, but excludes the 127 counties that never received railroad access. In this restricted sample, *log(years of rail+1)* no longer takes zero values and cleanly captures the intensive margin of railroad connection. The reduced form and 2SLS estimates closely resemble those in column 3. Column 6 omits the timing control in the same restricted sample and produces results similar to column 1. This confirms that the sign reversal is not driven by the exclusion of never-connected counties, but by properly accounting for the intensive margin of time since railroad connection.

Table 1 shows that residual variation in the instrument declines once both timing controls are included, raising potential concerns about multicollinearity between the instrument, the shift–share–type controls, and the railroad timing controls. Online Appendix Table A1 reports analogous results without the shift–share–type controls. Although this specification absorbs less variation in the instrument, the reduced form and 2SLS estimates still reverse sign, even as the first stage becomes very weak. SNQ also propose two alternative instruments to address concerns about correlation with railroad timing. Online Appendix A shows that these alternatives are subject to the same concern as the baseline instrument. Table Online Appendix A2 further demonstrates that the first stages become very weak once both the intensive and extensive margins of railroad access are controlled for and the alternative instruments thus do not provide sufficient identifying variation.

The results for the remaining outcomes in SNQ's Table 4 also turn insignificant or flip sign once including the binary indicator for never-connected counties, as shown in Online Appendix Table A3.[7][8] This issue is not confined to long-run outcomes. Online Appendix Table A5 shows that the positive effects on manufacturing output between 1860-1920 and in 1930 become weaker and lose statistical significance once both the extensive and intensive margins of railroad connection are controlled for.

4 Identification and inference

The re-analysis in Table 1 shows that SNQ's main results do not hold once the timing of railroad connection is controlled for. However, this specification does not address all

[6] I was unable to reproduce the F-statistics reported by SNQ. The F-statistic obtained through Conley (1999) standard errors in Column 1 (12.09) is considerably lower than reported by SNQ (21.222).

[7] This is unsurprising, since the other outcomes are strongly correlated with income per capita, with absolute correlation coefficients ranging from 0.44 to 0.77.

[8] Online Appendix Table A4 demonstrates that the null results for social outcomes reported in SNQ Table 5 are unaffected.

endogeneity concerns. Section 4.1 discusses identification and inference in SNQ's design when interpreted as a shift–share IV, and Section 4.2 re-analyzes the main results under alternative specifications.

4.1 SNQ's instrument as a shift–share IV

The recent literature on shift–share instruments offers a framework to clarify identification and inference (Borusyak, Hull and Jaravel, 2022; Goldsmith-Pinkham, Sorkin and Swift, 2020; Adao, Kolesár and Morales, 2019). This literature has shown that there are two pathways to identification with shift–share IVs: assuming that the shifts are exogenous (Borusyak, Hull and Jaravel, 2022), or assuming that the shares are exogenous (Goldsmith-Pinkham, Sorkin and Swift, 2020). As total migration to the US is unlikely to be endogenous to local county-level factors, the former seems plausible. SNQ also provide a compelling argument why the latter is implausible (p 392.): "counties that became connected to the railway network during certain periods ... may have disproportionately benefitted from being connected to the railway".

Borusyak, Hull and Jaravel (2022) show that exogenous shifts alone do not guarantee identification. The shares determine the degree of exposure to the decadal migration shocks. If the sum of shares differs across observations, some units are mechanically more exposed and therefore have more extreme instrument values. Put differently, the *expected* value of the instrument—averaging over possible realizations of the migration shocks—is proportional to the sum of shares. If the shares are not exogenous, this correlation between the instrument and the sum of shares implies a violation of the exclusion restriction. In the case of SNQ's instrument, the sum of shares can be written as:

$$\mu_i = \sum_{t=1860}^{1920} RR_{i,t-10} = \frac{R_i}{T_i} \tag{6}$$

R_i is the number of decades a county was connected to the railroad between 1860 and 1920, and T_i is the number of decades the county existed during this period. Similar to the instrument itself (Figure 1), the sum of shares is positively correlated with earlier railroad connection and negatively correlated with county age, as shown in Online Appendix Figure A5. Online Appendix Table A6 shows that the sum of shares is strongly correlated with the instrument, the average immigrant share, and log income per capita in 2000, even when baseline controls are included.

Unless additional assumptions are made about the exogeneity of the sum of shares, it should be controlled for to achieve identification. However, Borusyak, Hull and Jaravel (2025) note that this approach requires a sufficiently large number of shifts. They write: "it [the exogenous shifts approach] requires many shifts g_i ..., g_K. Otherwise, if K is a small number, the shifts may by chance be correlated with unobservables even if they are truly random". In the case of SNQ, there are only seven decadal shocks, so residual variation in the instrument may be correlated with unobserved confounders.

The small number of shifts also complicates inference, as a single decadal shock can exert substantial influence on the instrument values of many counties. Standard asymptotic inference does not account for the complex dependence structure of shift–share designs (Adao, Kolesár and Morales, 2019). Borusyak, Hull and Jaravel (2022) show that the 2SLS estimator can be re-expressed as a shift-level IV regression, allowing for valid inference using (cluster-)robust standard errors. In our setting, however, SNQ's instrument is constructed from fewer shocks than required covariates, rendering this approach infeasible.

In such cases, Borusyak and Hull (2023) recommend a randomization inference (RI) procedure using the Hodges-Lehmann T-statistic.[9] The RI procedure repeatedly draws shocks from a plausible counterfactual shock distribution, recomputes the instrument, and calculates the simulated test statistic. To assess how likely the original test statistic is to arise by chance, it is compared to the distribution of simulated test statistics. A natural counterfactual shock sequence is given by the 7! possible permutations of the seven realized decadal shocks across the seven decades. These sequences are in expectation uncorrelated to the realized shocks. By construction, these permutations are, in expectation, uncorrelated with the realized shock sequence.

4.2 Controlling flexibly for railroad timing and county age

Following Section 4.1, I re-estimate SNQ's baseline specification without explicit timing controls but include the sum of shares, μ_i. Column 1 of Table 2 shows that the first stage remains moderately strong after controlling for μ_i. At the same time, the 2SLS coefficient declines substantially in magnitude and becomes statistically insignificant at the 10% level. As discussed above, even when conditioning on μ_i, the instrument may still correlate with railroad timing or county age. Online Appendix Figures A3 and A4 show that this is indeed the case: both the residualized instrument and residualized income in 2000 are higher for earlier-connected and older counties.

I examine how results change when (additionally) controlling for the timing of railroad connection and county establishment in Online Appendix Table A7 and columns 2–4 of Table 2. Online Appendix Table A7 reports specifications that include linear controls for railroad timing and county age, with and without μ_i. A linear control for years since railroad connection—capturing the intensive margin only partially—yields estimates that are smaller and marginally significant. Once μ_i is also included, the coefficients are no longer statistically distinguishable from zero. If instead controlling linearly for county age, the estimate turns insignificant.

Columns 2–4 of Table 2 introduce saturated controls for the decade of railroad connection, decade of establishment, and both. Flexibly controlling for railroad timing again renders the 2SLS estimate negative, as in columns 3 and 4 of Table 1. Including decade of county establishment fixed effects in column 3 renders the first stage extremely weak.[10] Overall, these findings show that the conclusions depend strongly on the assumptions about whether and how county age and the timing of railroad connection affect long-run outcomes.

4.3 Randomization inference

To obtain valid inference, I follow Borusyak and Hull (2023) and perform randomization inference by permuting the seven decadal shocks across decades and recalculating the instrument. For each permutation, I calculate four test statistics: the Hodges-Lehmann statistic T for the 2SLS estimate as proposed by Borusyak and Hull (2023), the first stage t-statistic to assess the strength of the first stage; and the coefficient and the t-statistic of the reduced

[9] Online Appendix C provides additional details on the RI procedure.

[10] Online Appendix B discusses the variation underlying the first stage and why it becomes very weak when fixed effects for the decade of establishment are included.

Table 2: Controlling for μ_i and timing-related fixed effects

	(1)	(2)	(3)	(4)
Panel A: First stage	Average immigrant share, 1860–1920			
Instrument	5.587***	6.463***	1.391	3.053
	(1.914)	(2.544)	(1.358)	(2.535)
Panel B: Reduced Form	Log income per capita, 2000			
Instrument	7.414	-18.248**	7.035*	-17.045**
	(5.696)	(5.813)	(4.747)	(6.226)
Panel C: 2SLS	Log income per capita, 2000			
Average Immigrant	1.327	-2.823**	5.057	-5.583
Share 1860-1920	(1.147)	(1.348)	(6.019)	(4.722)
RI p-value	{0.763}	{0.284}	{0.867}	{0.334}
90% Anderson-Rubin CI	[-0.33,4.44]	[-8.1,-1.23]	$[-\infty,\infty]$	$[-\infty,\infty]$
Observations	2935	2935	2935	2935
First stage F	8.52	6.46	1.05	1.45
R^2 of IV on controls	0.970	0.980	0.958	0.986
Expected instrument μ_i	✓			
Decade of railroad connection FE		✓		✓
Decade of county establishment FE			✓	✓

Notes: First stage, reduced form and 2SLS estimates with alternative controls for timing of railroad connection and county age. All models include the baseline controls of Table 4 of SNQ, except for the *log(years of rail+1)* control. Column 1 controls for the sum of shares. Column 2 includes fixed effects for the decade of railroad connection (including a level for never connected counties), column 3 includes fixed effects for the decade of county establishment and column 4 includes both. Conley standard errors using a five-degree window are shown in parentheses. Anderson-Rubin confidence intervals (square brackets) and Randomization Inference p-values (curly brackets) are reported for the key 2SLS coefficient of interest. Anderson Rubin confidence interval bounds are evaluated to $\pm\infty$ if ± 15 are not included in the confidence interval. Kleibergen-Paap F-statistics based on Conley standard errors are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

form model. For each specification and test statistic, I plot the distribution across permutations and indicate the corresponding test statistic from the original instrument with a dashed vertical line. Randomization inference p-values are calculated as the fraction of simulated statistics that are at least as extreme as the observed statistic. Additional details are provided in Online Appendix C.

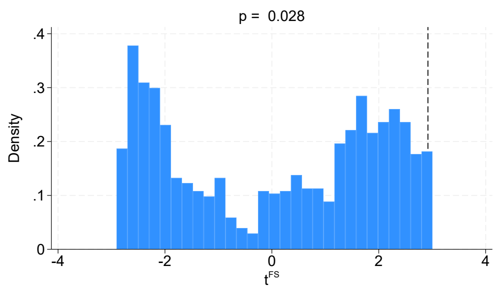
Online Appendix Figure A7 reports the results for SNQ's baseline specification. The realized 2SLS T-statistic is not extreme relative to its permutation distribution. Randomization inference therefore leads to markedly different conclusions compared with the conventional asymptotic inference based on Conley standard errors used by SNQ: while the conventional p-value is below 0.01, the RI p-value is 0.791. The reduced form t-statistics are highly asymmetric and skewed toward positive values. 53% of t-statistics exceed 2, implying that for randomly reassigned shocks, conventional inference frequently rejects the null in favor of a positive effect of historical migration on log income per capita in 2000.

Controlling for μ_i partially mitigates these issues. Figure 4 shows that for the specification of column 1 of Table 2 the test statistic distributions become more symmetric and centered around zero. The first-stage t-statistic is relatively large (p=0.028), but the T-statistic for the 2SLS estimate is not extreme at all (p=0.763).[11] The reduced form results further illustrate the shortcomings of asymptotic inference for this specification: 61% of simulated t-statistics lie below -2 or above 2.

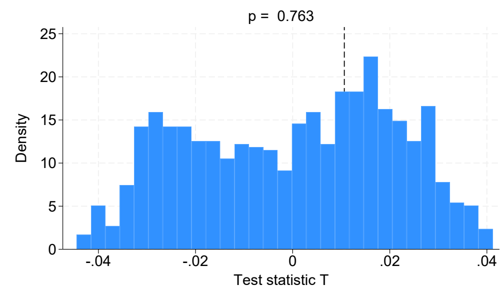
[11] Online Appendix Figures A8 and A9 show that sampling shocks with replacement or from a normal distribution, with the same mean and standard deviation as the realized shocks, leads to similar conclusions.

Tables 1 and 2 report RI p-values based on the Hodges-Lehmann T-statistic across specifications; these are consistently much larger than conventional significance levels. The RI framework also allows for the construction of valid confidence intervals by inverting the test on the Hodges-Lehmann T-statistic for different candidate parameter values. Online Appendix Figure A10 shows that the 90% confidence interval of the effect of a 1 percentage point higher migration share during the Age of Mass Migration on per capita income in 2000 ranges from a reduction of 12% to an increase of 8%, which is uninformatively large.

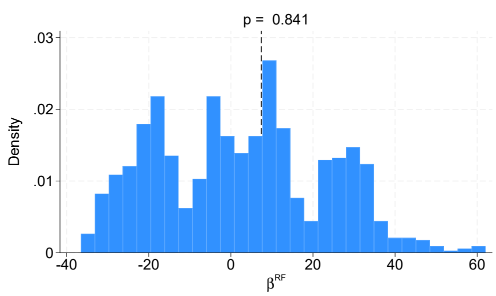
Figure 4: Distribution of RI test statistics by permuting decadal shocks, controlling for μ_i



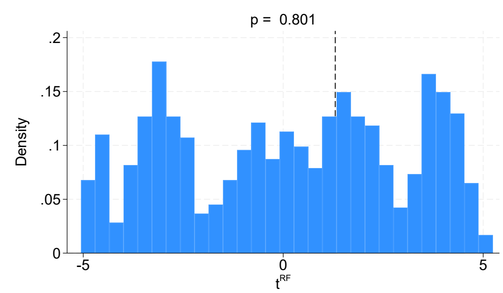
(a) First-stage t-statistic



(b) Hodges-Lehmann T-statistic



(c) Reduced form coefficient



(d) Reduced form t-statistic

Notes: Distribution of Randomization Inference test statistics, based on a specification that includes all baseline controls as well as the sum of shares μ_i, but excludes the control for *log(years of rail+1)*. Shocks are drawn from the set of 7! permutations of the seven decadal shocks across seven decades. (a) shows the t-statistic of the coefficient on the instrument in the first stage. (b) shows the distribution of the Hodges-Lehmann statistic T, given by T= $\frac{1}{N}\sum_i(z_i - \theta\mu_i)(y_i - bx_i)$, where θ is the mean of the seven decadal shocks. The null hypothesis is that $b = 0$. (c) shows the distribution of coefficient estimates on the instrument from the reduced form regression. (d) shows the distribution of t-statistics of the coefficient on the instrument in the reduced form regression. Dashed lines indicate the value of the respective statistic for the realized shock. t-statistics are calculated using Conley standard errors adjusted for spatial correlation using a five-degree window. Randomization Inference p-values are defined as the share of simulations that yield a more extreme than the realized statistic, times 2. Based on N = 999 randomly drawn permutations.

5 Conclusion

This comment shows that SNQ's core results disappear once the timing of railroad connection or county age is properly controlled for. This matters because both variables (i) are mechanically related to county-level migration during the Age of Mass Migration and (ii) are plausibly linked to long-run economic outcomes through channels other than migration. The analysis also indicates that the identifying variation based on the seven decadal migration shocks is too limited to deliver precise estimates, regardless of exact identification assumptions.

How migration during the Age of Mass Migration affected long-run prosperity in the US therefore remains an open empirical question. Future research must rely on stronger sources of exogenous variation in historical settlement patterns. Recent contributions that uncover new sources of plausibly exogenous variation may offer viable and sufficiently powerful alternatives to the railroad-based instrument (Burchardi, Chaney and Hassan, 2019; Obolensky, Tabellini and Taylor, 2024). Another promising direction is to exploit the geography of migrant entry points together with the network structure of railroads to capture variation in travel costs and predict settlement patterns across counties. This is left for further research.

References

Adao, Rodrigo, Michal Kolesár, and Eduardo Morales. 2019. "Shift-share designs: Theory and inference." *Quarterly Journal of Economics*, 134(4): 1949–2010.

Borusyak, Kirill, and Peter Hull. 2023. "Nonrandom exposure to exogenous shocks." *Econometrica*, 91(6): 2155–2185.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2022. "Quasi-experimental shift-share research designs." *Review of Economic Studies*, 89(1): 181–213.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2025. "A practical guide to shift-share instruments." *Journal of Economic Perspectives*, 39(1): 181–204.

Burchardi, Konrad B, Thomas Chaney, and Tarek A Hassan. 2019. "Migrants, ancestors, and foreign investments." *Review of Economic Studies*, 86(4): 1448–1486.

Conley, Timothy G. 1999. "GMM estimation with cross-sectional dependence." *Journal of Econometrics*, 92(1): 1–45.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik instruments: What, when, why, and how." *American Economic Review*, 110(8): 2586–2624.

Obolensky, Marguerite, Marco Tabellini, and Charles Taylor. 2024. "Homeward bound: How migrants seek out familiar climates." Unpublished working paper.

Sequeira, Sandra, Nathan Nunn, and Nancy Qian. 2020. "Immigrants and the making of America." *Review of Economic Studies*, 87(1): 382–419.

Online Appendix

(For online publication only)

A Alternative instruments

SNQ employ several alternative instrumental variables to reduce endogeneity concerns. The first uses weather shocks to predict decadal migration from Europe. This design reduces endogeneity concerns related to European migration reacting to railroad connection in counties with large growth potential, but does not address the issue outlined in Section 3.1.

SNQ also introduce two alternative instruments in column 2 and 3 of their Table 11 motivated by concerns about the correlation between the original instrument and timing of railroad connection. When doing so, the authors include the never-connected dummy but remove the *log(years of rail+1)* control in these specifications, without providing a rationale for the latter.[12] Hence, also these specifications do not control for time since railroad connection.

The first alternative instrument simply disregards periods without railway connection and averages migration shocks across all decades after receiving a railroad connection. This instrument disregards the variation driven by differential length of exposure to the railroad network and is thus unsurprisingly very weak.[13] The second alternative instrument only considers migration flows in the decade following the decade of railroad connection, which yields a value of zero for counties connected in the 1920s and never-connected counties.

Both alternative instruments can be written in the form of equation 4, with the same shifts but different shares. The share part of the first alternative instrument can be written as:

$$RR_{i,t-10}^{avg.connected} = \begin{cases} \dfrac{1}{N_i^{RR}}, & \text{if } I_{i,t-10}^{\text{RR Access}} = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

SNQ here define N_i^{RR} as the number of *years* a cohort was connected to the railways between 1830 and 1920.[14] Let R_i^0 denote the decade in which a county received railroad connection. The sum of shares for this instrument is given by:

$$\frac{\min(1920 - R_i^0, 70)/10}{(1930 - R_i^0)} \tag{8}$$

For example, for those first connected to railroads in the 1910s, the sum of shares is $\frac{1}{20}$. For those counties never connected or first connected in the 1920s the sum of shares is equal to 0.

[12] SNQ note (footnote 57): *"Since the predicted average immigrant share instrument for counties that are never connected to the railway network is zero, the specifications include an indicator variable for whether the county was never connected to the railway."* However, this issue is not unique to the alternative instruments, but also arises for the original instrument.

[13] Column 2 of Table 11 of SNQ reports the first-stage F-statistic of 18.803. However, as with the F-statistics in Table 4 of SNQ, I cannot reproduce this: the F-statistic based on Conley standard errors with a 5 degree window is only 3.7.

[14] In their paper SNQ write (p. 412): where "N_i^{RR} is the number of time periods". However, in the code it is defined as the number of years, which I use in this section.

The share part of the second alternative instrument can be written as:

$$RR^{first}_{i,t-10} = \begin{cases} 1, & \text{if } I^{\text{RR Access}}_{i,t-10} = 1 \text{ and } I^{\text{RR Access}}_{i,t-20} = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Hence, the sum of shares is equal to 1 for counties connected in 1910 or earlier and 0 for counties connected in the 1920s or later. Although those alternative instruments are intended to not be correlated to the timing of railroad connection, the exposure to shocks is not equal across railroad cohorts and the realization still depends on the realization of the seven decadal migration shocks. To illustrate that these alternatives also have lower instrument values for later decades, I plot the average instrument values by decade in Online Appendix Figure A2b, analogous to Online Appendix Figure A2a. This shows that the alternative instruments are still negatively correlated between the 1830 and the 1910 railroad cohorts. Online Appendix Figure A11 shows that both alternative instruments also exhibit a residual correlation to timing of railroad connection. Hence, the concern about upward bias also applies to these alternative instruments. Online Appendix Table A2 shows that when including both the *log(years of rail+1)* and the never connected control, the first stages for both instruments turn very weak.

B Why does the first stage vanish when controlling for decade of establishment FE?

Table 2 and Online Appendix Table A7 show that the instrument remains relevant when controlling (flexibly) for the timing of railroad connection and when adding a linear control for county age. However, in columns 3 and 4 of Table 2 the first stage loses relevance once saturated controls for the decade of county establishment are included.

Why does this happen? Online Appendix Figure A12 shows that the residualized average immigrant share follows an inverse-U pattern across decades of establishment. Counties established during the early mass migration period (1860–1880) exhibit substantially higher immigrant shares (3–12 percentage points higher), while both older and younger counties display lower shares. The residualized instrument in Online Appendix Figure A4a mirrors this pattern: counties founded between 1860 and 1890 have positive residuals after accounting for the timing of railroad connection. This coinciding establishment cohort-specific variation explains part of the first stage effect.

The pattern is plausible. Counties founded at the onset of mass migration may have attracted more immigrants than long-established counties or those not yet formed as labor demand may be particularly high in newly created counties and migrants are more willing to settle there than native-born. SNQ do not discuss this and may not have intended to rely on variation in the instrument that depends on the age of counties.

Whether this variation is desirable is less clear. Establishment cohort-specific factors may threaten the exclusion restriction. Older (often coastal) counties may still be richer today due to persistent geographic advantages. Conversely, some younger counties may have successfully attracted new industries.

C Randomization inference

This Appendix discusses the Randomization Inference procedure introduced in Section 4.2, which closely follows Borusyak and Hull (2023). The core idea is to re-draw the seven decadal shocks from a distribution of plausible counterfactual shocks, recalculate the instrument, rerun the regressions, calculate relevant test statistics and compare the original test statistic to the distribution.

The first step concerns specifying realistic counterfactual shocks. I specify the following three shock generation processes:

1. Randomly permuting the realized shocks. This is equivalent to re-drawing without replacement. This yields 7! possible permutations.

2. Drawing shocks from the set of realized shocks, with replacement. This yields 7^7 possible permutations.

3. Drawing shocks from a normal distribution associated with the mean and standard deviation of the realized distribution of shocks, as shown in Online Appendix Figure A6.

I draw $N_b = 999$ instances of these shocks, recalculate the instrument and calculate the following test statistics for each simulated instrument (as well as the original instrument):

- First-stage t-statistic
  To assess how strong the original first stage is compared to the simulated first stages, I calculate the first-stage t-statistic.

- Hodges-Lehmann T-statistic
  Borusyak and Hull (2023) recommend to calculate the Hodges-Lehmann T-statistic to assess the strength of the 2SLS evidence: T$= \frac{1}{N}\sum_i (Z_i - \theta\mu_i)(y_i - bx_i)$, where Z_i is the (recalculated) instrument, θ is the mean of the decadal migration shocks, μ_i the sum of shares and b^{cand} is the candidate parameter value. $\theta\mu_i$ is the expected value of the instrument. We are interested whether $b = 0$. Hence, T is equivalent to the covariance between the "recentered" instrument and the outcome. Note that recentering (subtracting the expected instrument $\theta\mu_i$) yields a secular shift independent of the realization of the shocks and instrument.

- Reduced form coefficient β^{RF}
  To assess how large the realized reduced form coefficient is relative to the coefficient obtained from randomly drawn shocks, I calculate the reduced form coefficient.

- Reduced form t-statistic τ^{RF}
  Compared to the reduced form β, the t-statistic has the advantage that it is also informative of the strength of the evidence, not solely the size of the coefficient.

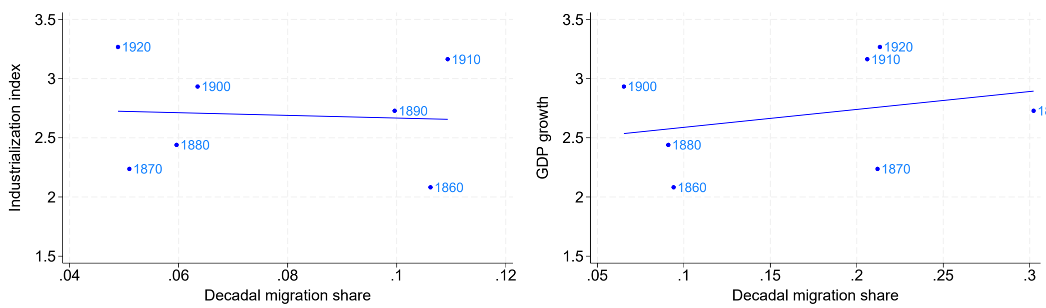
After calculation of these test statistics, I compare how extreme the test statistics based on the actual instrument are in comparison to those from the simulated instruments. Specifically, for each test statistic I calculate two-sided Randomization Inference p-values in the following way:

$$\hat{p}_{\text{RI}} = \frac{2}{N_b} \min\left(\sum_{n=1}^{N_b} \mathbf{1}\left\{ T^{(n)} \geq T^{\text{orig}} \right\}, \sum_{n=1}^{N_b} \mathbf{1}\left\{ T^{(n)} \leq T^{\text{orig}} \right\} \right) \tag{10}$$

Here, $n = \{1, ..., N_b\}$ indexes the simulations, and *orig* indicates that it concerns the original realized instrument. T's indicate either of the four test statistics described above. This test can also be used to construct valid confidence intervals through inversion of the test described above for different candidate values b^{cand}. For each b^{cand}, I calculate the randomization inference p-value. If the p-value exceeds α, the candidate b^{cand} falls inside the $(1-\alpha)\%$ confidence interval. If the p-value is smaller than α, the candidate b^{cand} falls outside the confidence interval or set.

D Additional Figures

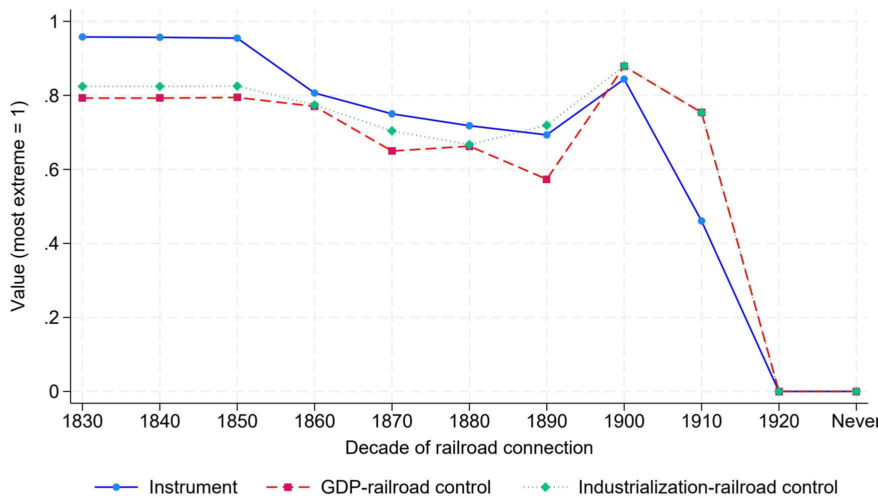
Figure A1: Correlation of decadal shocks: industrialization and GDP growth vs. migration



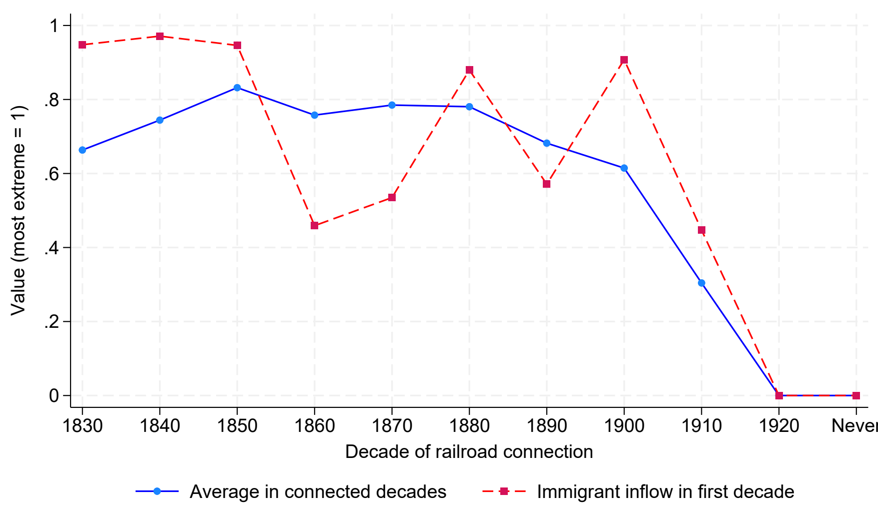
(a) Industrialization (b) GDP growth

Notes: Correlation of the seven decadal migration shocks with (a) the industrialization index ($\rho = -0.07$) and (b) the GDP growth ($\rho = +0.17$).

Figure A2: Average instrument control function values by decade of railroad connection.



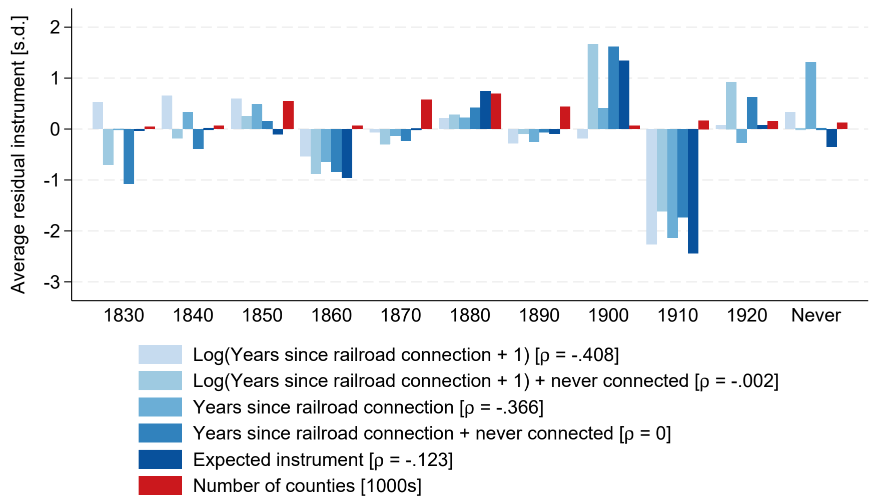
(a) Average instrument and industrialization and GDP growth control functions



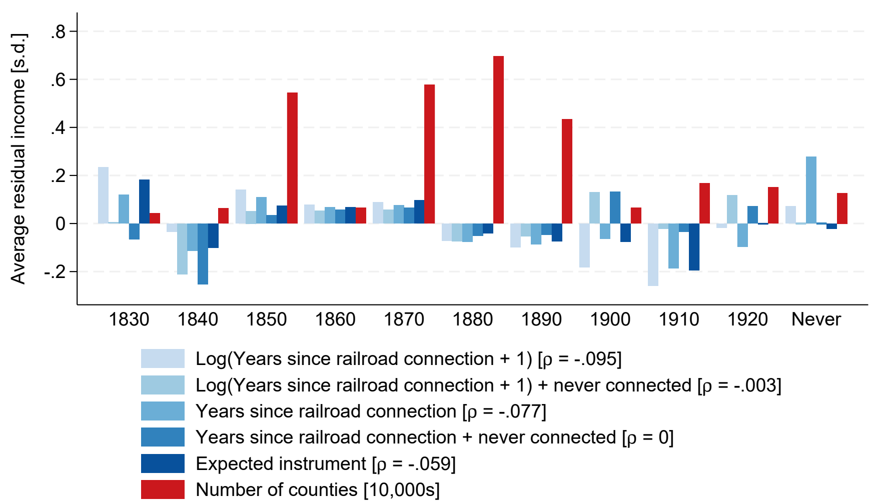
(b) Reduced form coefficient

Notes: a) Average instrument and shift–share-like industrialization and GDP growth control function values by decade of railroad connection. b) Average alternative instrument values used in SNQ Table 11 by decade of railroad connection. For visualization purposes I divide each time series by its most extreme value across all counties and decade of railroad connection. N = 2,935.

Figure A3: Residualized instrument and residualized log income per capita by decade of railroad connection



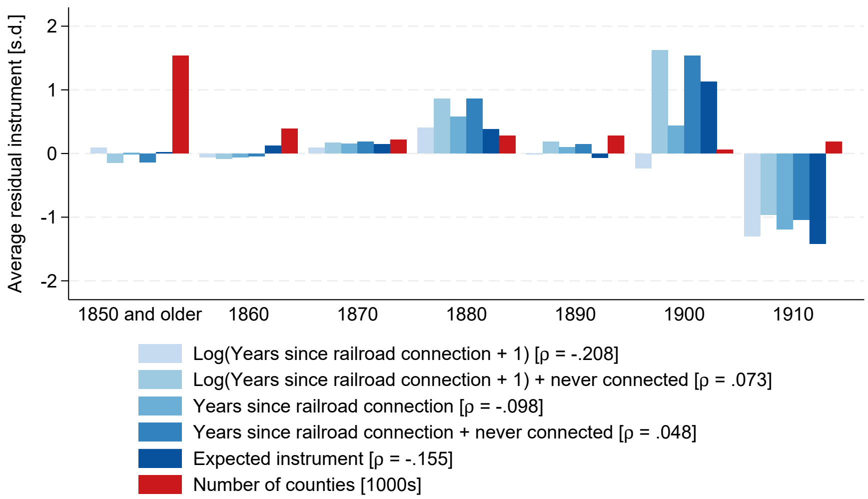
(a) Instrument



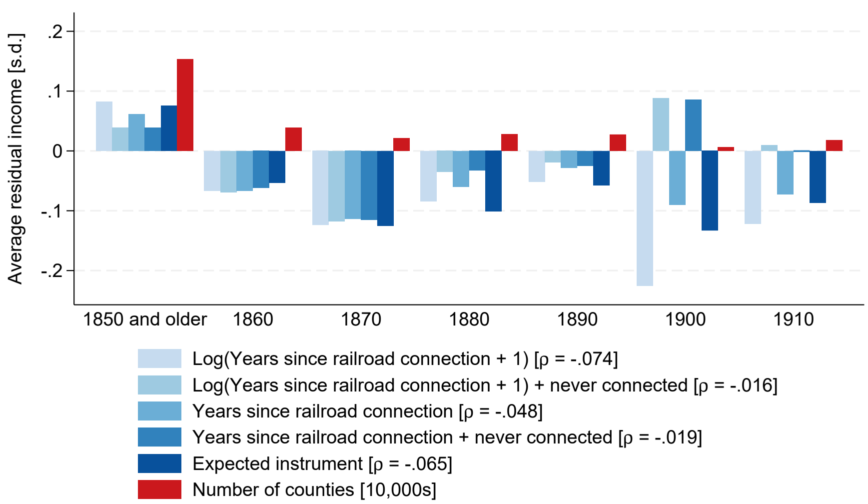
(b) Log income per capita in 2000

Notes: Average residualized and standardized a) instrument and b) log income per capita in 2000 by decade of railroad connection, for different sets of controls. Each specification includes the baseline controls of Table 4 of SNQ (excluding the control for *log(years of rail+1)*, unless noted otherwise) and the controls mentioned in the legend. The legend reports the correlation between the residualized instrument and the decade of railroad connection, on the sample of ever-connected counties. The number of observations by cohort are indicated by the red bars.

Figure A4: Residualized instrument and residualized log income per capita in 2000 by decade of establishment



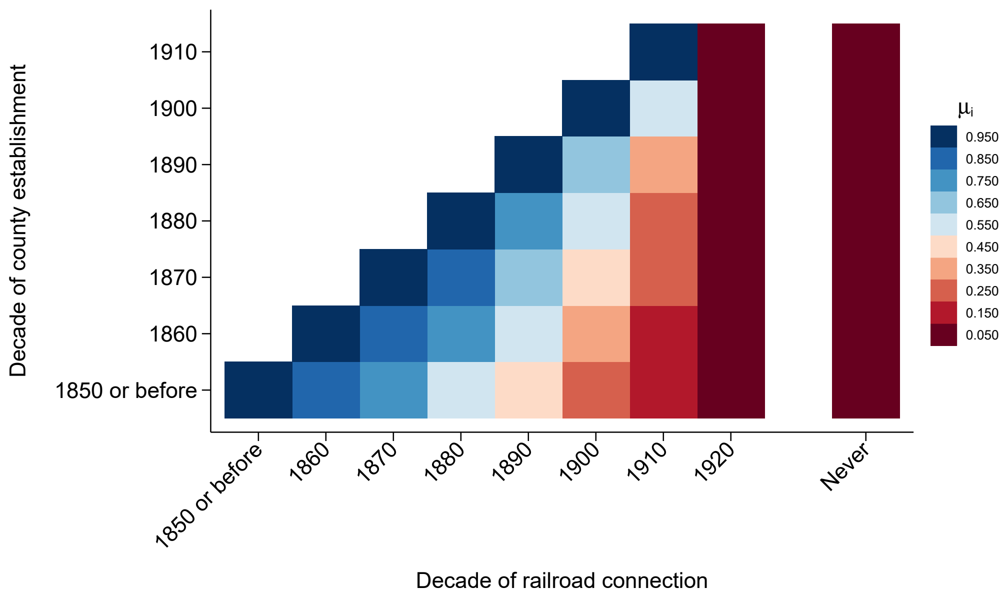
(a) Instrument



(b) Log income per capita in 2000

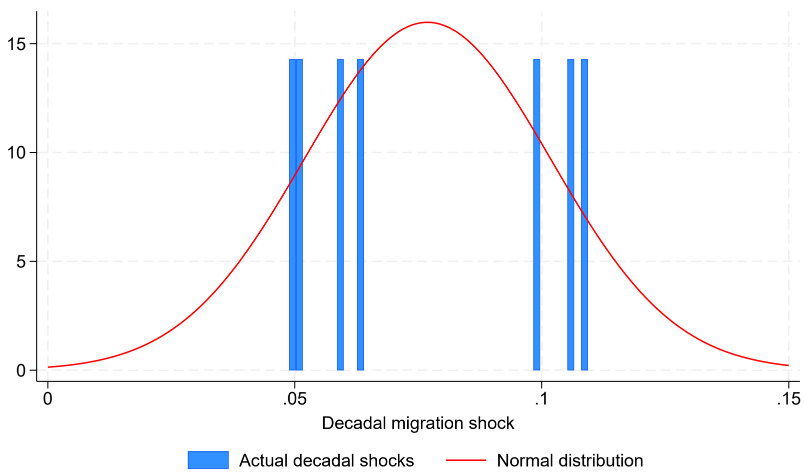
Notes: Average residualized and standardized a) instrument and b) log income per capita in 2000 by decade of establishment, for different sets of controls. Each specification includes the baseline controls of Table 4 of SNQ (excluding the control for *log(years of rail+1)*, unless noted otherwise) and the controls mentioned in the legend. The legend reports the correlation between the residualized instrument and the decade of establishment. The number of observations by cohort are indicated by the red bars.

Figure A5: Sum of shares μ_i as a function of establishment and railroad cohorts



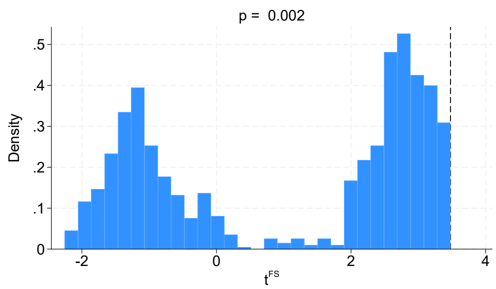
Notes: Sum of shares μ_i by decade of establishment and decade of railroad connection. N = 2,935.

Figure A6: Histogram of decadal migration shocks and implied normal distribution

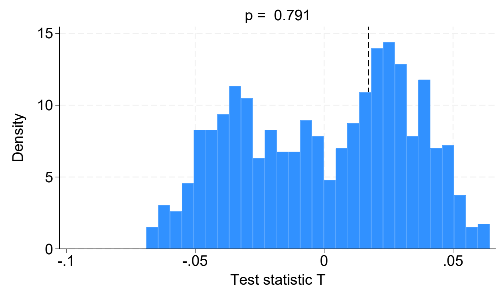


Notes: Histogram of the seven decadal shocks and associated normal distribution with mean 0.076 and sd 0.025.

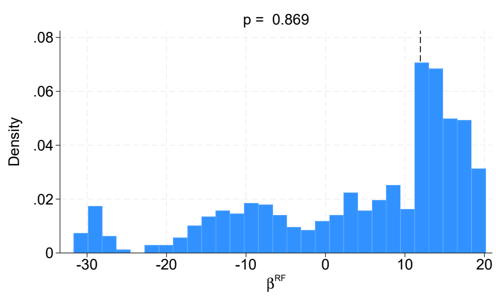
Figure A7: Distribution of RI test statistics by permuting decadal shocks, original specification



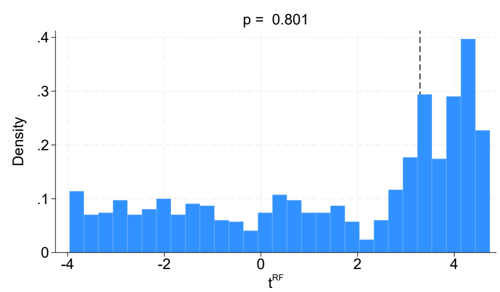
(a) First-stage t-statistic



(b) Hodges-Lehmann T-statistic



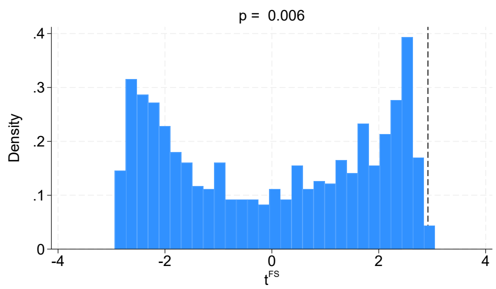
(c) Reduced form coefficient



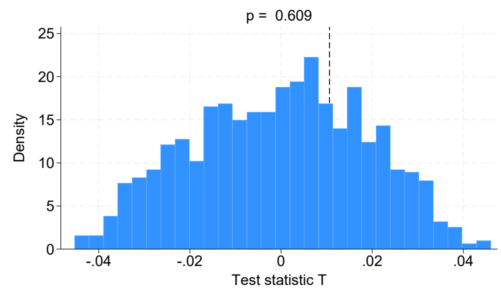
(d) Reduced form t-statistic

Notes: Distribution of Randomization Inference test statistics, based on the same specification as Table 4 of SNQ. Shocks are drawn from the set of permutations of the seven decadal shocks, without replacement. For details on the procedure, see notes to Figure 4. Based on N = 999 permutations.

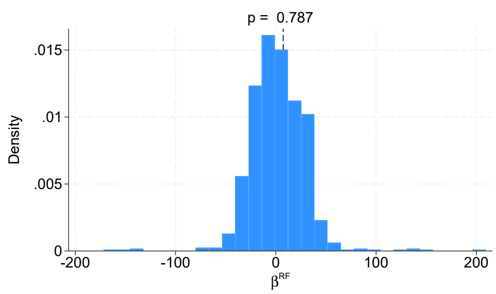
Figure A8: Distribution of RI test statistics by redrawing shocks with replacement, controlling for μ_i



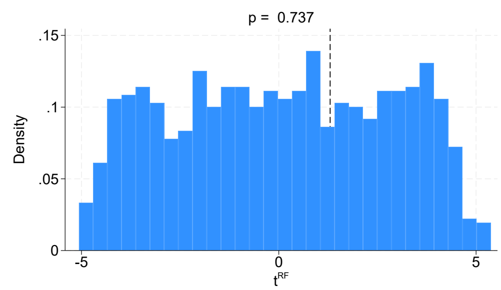
(a) First-stage t-statistic



(b) Hodges-Lehmann T-statistic



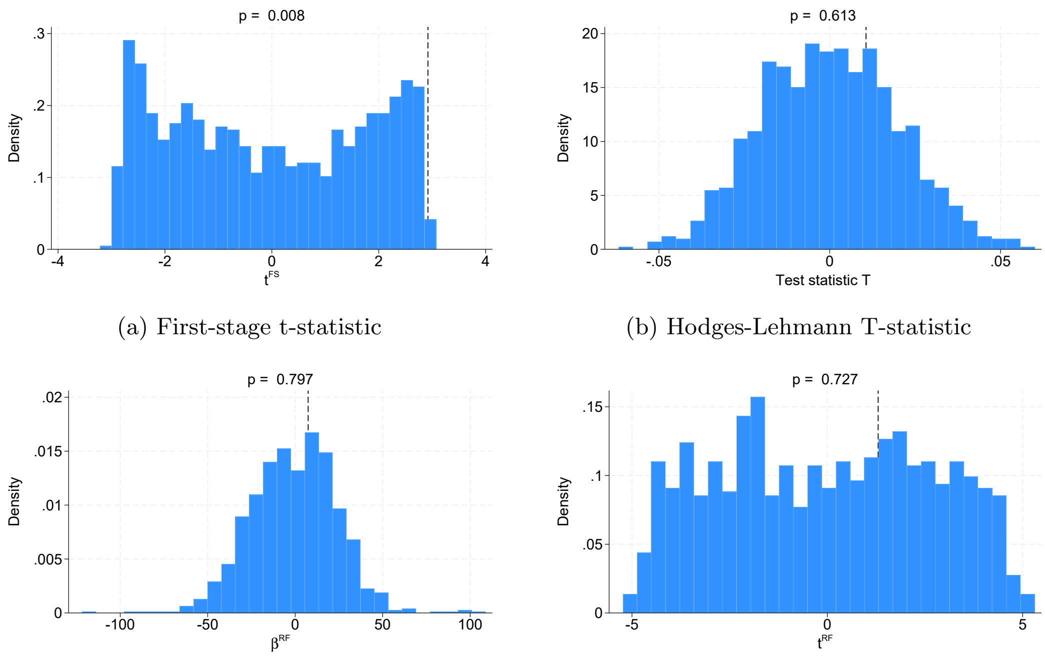
(c) Reduced form coefficient



(d) Reduced form t-statistic

Notes: Distribution of Randomization Inference test statistics, based on a specification that includes the sum of shares μ_i but excludes the control for *log(years of rail+1)*. Shocks are drawn from the set of permutations of the seven decadal shocks, with replacement. For details on the procedure, see notes to Figure 4. Based on N = 999 permutations.

Figure A9: Distribution of RI test statistics by drawing from normal distribution, controlling for μ_i



(a) First-stage t-statistic

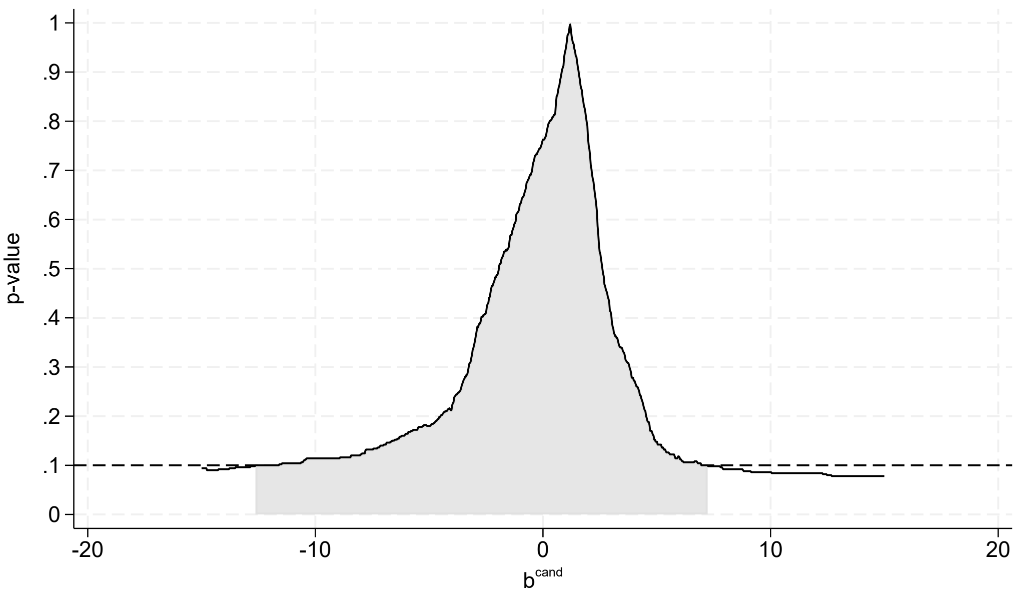
(b) Hodges-Lehmann T-statistic

(c) Reduced form coefficient

(d) Reduced form t-statistic

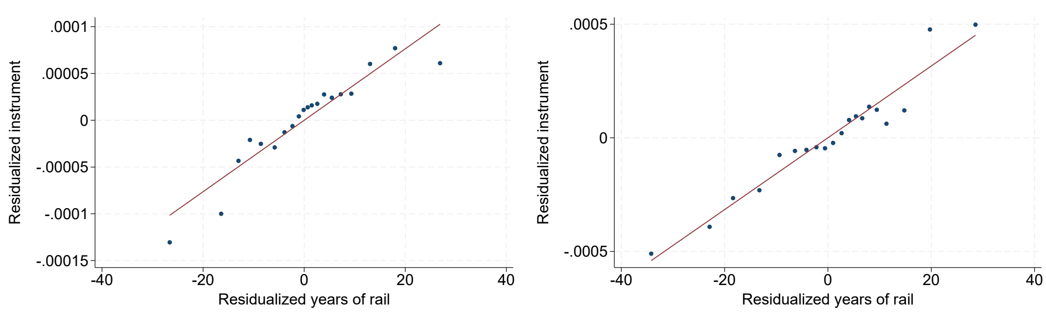
Notes: Distribution of Randomization Inference test statistics, based on a specification that includes the sum of shares μ_i but excludes the control for *log(years of rail+1)*. Shocks are drawn from a normal distribution with a mean and standard deviation equal to that of the realized shocks. For details on the procedure, see notes to Figure 4. Based on N = 999 permutations.

Figure A10: Randomization Inference Confidence Intervals



Notes: Randomization inference p-values for different candidate values b^{cand} based on the Hodges-Lehmann T-statistic as suggested by Borusyak and Hull (2023). The specification includes all baseline controls (excluding the *log(years of rail+1)*) and the sum of shares. The instrument is recalculated by randomly permuting the shocks across decades 999 times. The p-value for every b^{cand} is obtained according to the procedure discussed in Online Appendix C for every b^{cand} between -15 and +15 with steps of 0.3. The area shaded in gray shows the area where the test does not reject the null using $\alpha = 0.1$. The implied 90% confidence interval ranges from -13.2 to 7.5. The corresponding effect size of a one percentage point higher migration share ranges from -12% to +8%.

Figure A11: Binned scatterplot of the residualized alternative instruments and time since railroad connection

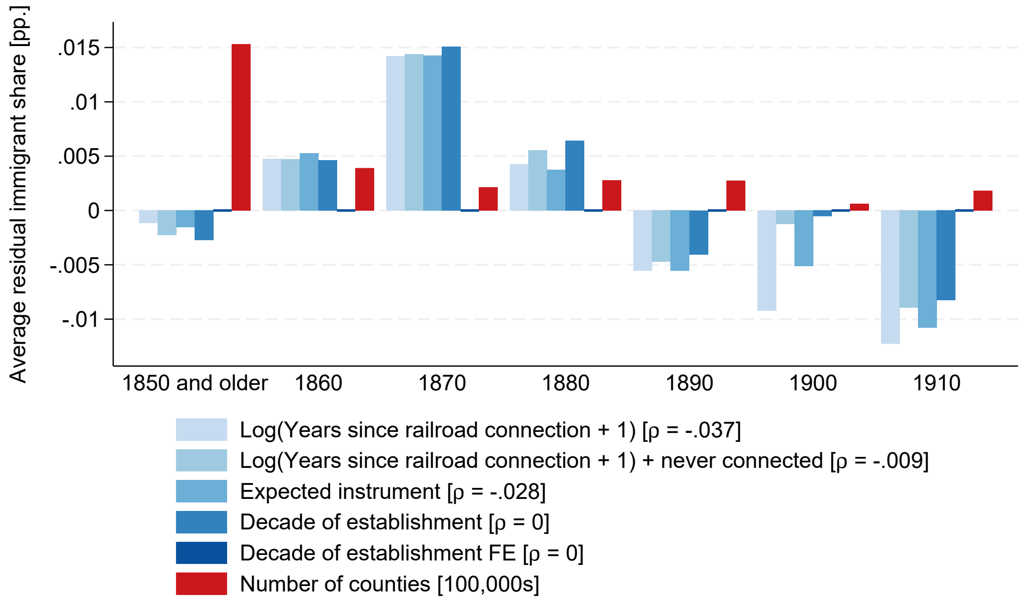


(a) Average over connected decades

(b) First connected decade

Notes: Binned scatterplot ($N_{bin} = 20$) with trendline of the residualized alternative instruments and the residualized number of years since railroad connection. The instruments are (a) the average migration shock over connected decades and (b) the migration shock in the first decade following railroad connection. The control variables used for residualization are the same as columns 2 and 3 of Table 11 of SNQ, which include a never-connected dummy but do not include the control for *log(years of rail+1)*. N = 2,935.

Figure A12: Residualized average immigrant share (1860-1920) by decade of county establishment for different controls



Notes: Average residualized immigrant share between 1860 and 1920 by decade of county establishment, for different sets of controls. Each specification includes the baseline controls of Table 4 of SNQ (excluding the control for the *log(years of rail+1)*, unless noted otherwise) and the controls mentioned in the legend. The legend also reports the correlation between the average immigrant share (1860-1920) and the decade of establishment. The number of observations by cohort are indicated by the red bars.

E Additional Tables

Table A1: Table 1, excluding the shift-share-like control functions

Dependent variable:	Log average per capita income in 2000					
	Full sample				Ever connected by rail	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: First stage	Average immigrant share, 1860–1920					
Instrument	1.590***	1.584***	0.635	0.676	0.493	1.516***
	(0.479)	(0.458)	(0.844)	(0.775)	(0.796)	(0.456)
Panel B: Reduced Form	Log income per capita, 2000					
Instrument	6.679***	7.146***	-2.799	-2.029	-3.604**	7.304***
	(1.695)	(1.707)	(2.130)	(2.090)	(1.968)	(1.812)
Panel C: 2SLS	Log income per capita, 2000					
Average Immigrant Share 1860-1920	4.201***	4.512***	-4.407	-3.003	-7.316	4.817***
	(1.734)	(1.813)	(6.257)	(4.169)	(12.656)	(2.024)
Observations	2935	2935	2935	2935	2808	2808
First stage F-statistic	11.02	11.95	0.57	0.76	0.38	11.06
R^2 of IV on controls	0.507	0.431	0.796	0.770	0.735	0.246
Baseline controls	✓	✓	✓	✓	✓	✓
Log(years of rail + 1)	✓		✓		✓	
Never connected		✓	✓	✓		
Years of rail				✓		

Notes: First stage, reduced form and 2SLS estimates based on column 1 of Table 4 of SNQ. All models include the baseline controls of Table 4 of SNQ, except for the log(years of rail+1) and the two shift-share-like control functions for industrialization and GDP growth. Column 1 presents the original estimates and columns 2-4 introduce different sets of control variables. Column 5 and 6 exclude the 127 never-connected counties; column 5 includes the baseline controls, whereas column 6 removes the control for log(years of rail + 1). Coefficients on additional controls are not shown in Panel A for brevity. Conley standard errors using a five-degree window are shown in parentheses. Kleibergen-Paap F-statistics based on Conley standard errors are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Alternative instrument: controlling for log (x+1) and never-connected dummy

	Average in connected decades			Immigrant inflow in first-connected decade		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Log income per capita, 2000					
Average Immigrant Share 1860-1920	8.532**	8.172**	48.806	4.540***	4.179***	-1.926
	(4.497)	(4.204)	(2719.172)	(1.773)	(1.686)	(4.254)
Observations	2935	2935	2935	2935	2935	2935
First stage F-statistic	3.70	3.92	0.00	11.00	10.02	0.68
R^2 of IV on controls	0.659	0.789	0.789	0.960	0.759	0.978
Baseline controls	✓	✓	✓	✓	✓	✓
Log(years of rail + 1)		✓	✓		✓	✓
Never connected	✓		✓	✓		✓

2SLS estimates based on the two alternative instruments: the average migrant share in all connected decades (columns 1-3) and the migrant share in then first decade after connection (columns 4-6). All models include the baseline controls of Table 4 of SNQ. columns 2 and 5 additionally control for the log(years of rail + 1) control and omits the binary indicator for never-connected counties. Columns 3 and 6 controls for both. Conley standard errors using a five-degree window are shown in parentheses. Kleibergen-Paap F-statistics based on Conley standard errors are shown. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Re-analysis of SNQ Table 4

	Log income per capita in 2000		Share below poverty line		Unemployment rate		Urbanization rate		Average years of schooling	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: OLS										
Average Immigrant Share 1860-1920	0.243***	0.211***	0.015	0.022	0.020*	0.024**	0.949***	0.913***	0.020	-0.080
	(0.130)	(0.130)	(0.028)	(0.028)	(0.015)	(0.015)	(0.184)	(0.189)	(0.307)	(0.306)
Panel B: Reduced Form										
Instrument	11.942***	-7.767*	-2.229***	2.228**	-1.876***	0.010	22.382***	5.655	41.925***	-11.143
	(3.629)	(4.474)	(0.777)	(1.308)	(0.500)	(0.639)	(6.820)	(9.774)	(10.562)	(13.148)
Panel C: 2SLS										
Average Immigrant Share 1860-1920	2.619***	-1.977	-0.489***	0.567*	-0.411***	0.002	4.909***	1.439	9.195***	-2.836
	(1.022)	(1.452)	(0.209)	(0.409)	(0.151)	(0.162)	(2.008)	(2.584)	(3.392)	(3.488)
Observations	2935	2935	2935	2935	2935	2935	2935	2935	2935	2935
First stage F-statistic	12.09	4.71	12.09	4.71	12.09	4.71	12.09	4.71	12.09	4.71
log(years of rail+1)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Never connected		✓		✓		✓		✓		✓

Notes: OLS, reduced form and 2SLS estimates of Table 4 of SNQ. Odd columns control for all baseline controls including the *log(years of rail+1)* control, even columns additionally control for a dummy for never connected counties. Conley standard errors using a five-degree window are shown in parentheses. Kleibergen-Paap F-statistics based on Conley standard errors are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Re-analysis of SNQ Table 5

	Social capital		Voting turnout		Total crime rate		Crimes against persons		Crimes against property	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: OLS										
Average Immigrant Share 1860-1920	-0.048**	-0.047**	-0.071**	-0.073**	0.008***	0.008***	0.002***	0.002***	0.004***	0.004***
	(0.029)	(0.030)	(0.046)	(0.046)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Panel B: Reduced Form										
Instrument	0.210	1.441	1.244	1.471	0.086	0.055	0.020	0.012	0.054	0.031
	(0.958)	(1.192)	(1.662)	(2.060)	(0.070)	(0.108)	(0.013)	(0.022)	(0.053)	(0.075)
Panel C: 2SLS										
Average Immigrant Share 1860-1920	0.046	0.364	0.271	0.378	0.019	0.014	0.004	0.003	0.012	0.008
	(0.209)	(0.326)	(0.347)	(0.517)	(0.017)	(0.028)	(0.003)	(0.005)	(0.012)	(0.019)
Observations	2934	2934	2925	2925	2935	2935	2935	2935	2935	2935
First stage F-statistic	12.09	4.71	12.09	4.71	12.09	4.71	12.09	4.71	12.09	4.71
log(years of rail+1)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Never connected		✓		✓		✓		✓		✓

Notes: OLS, reduced form and 2SLS estimates of Table 5 of SNQ. Odd columns control for all baseline controls including the *log(years of rail+1)* control, even columns additionally control for a dummy for never connected counties. Conley standard errors using a five-degree window are shown in parentheses. Kleibergen-Paap F-statistics based on Conley standard errors are reported.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Re-analysis of SNQ Table 6

	Log Average Manufacturing Output per Capita				Log Average Manufacturing Output per Establishment				Log Number Establishments per 10,000 Inhabitants			
	1860-1920		1930		1860-1920		1930		1860-1920		1930	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: OLS												
Average Immigrant Share 1860-1920	3.476***	3.423***	4.216***	4.094***	3.301***	3.275***	3.343***	3.222***	0.319**	0.263*	0.783***	0.782***
	(0.631)	(0.638)	(0.796)	(0.802)	(0.537)	(0.540)	(0.648)	(0.653)	(0.249)	(0.250)	(0.248)	(0.249)
Panel B: Reduced Form												
Instrument	40.765*	13.672	74.736***	-4.844	20.778	0.042	71.924***	-4.705	32.710***	9.121	2.079	-0.204
	(33.988)	(41.084)	(26.368)	(42.989)	(29.227)	(35.276)	(23.653)	(38.610)	(6.462)	(10.231)	(6.765)	(9.776)
Panel C: 2SLS												
Average Immigrant Share 1860-1920	9.014*	3.487	16.197***	-1.116	4.594	0.011	15.588***	-1.084	7.253***	2.315	0.453	-0.047
	(8.460)	(11.195)	(7.343)	(9.821)	(6.838)	(8.998)	(6.868)	(8.800)	(2.389)	(2.859)	(1.467)	(2.400)
Observations	2805	2805	2463	2463	2805	2805	2463	2463	2804	2804	2462	2462
First-stage F-statistic	6.99	1.97	6.01	2.22	6.99	1.97	6.01	2.22	6.99	1.97	6.01	2.22
log(years of rail+1)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Never connected		✓		✓		✓		✓		✓		✓

Notes: OLS, reduced form and 2SLS estimates of Table 6 of SNQ. Odd columns control for all baseline controls including the *log(years of rail+1)* control, even columns additionally control for a dummy for never connected counties. Conley standard errors using a five-degree window are shown in parentheses. Kleibergen-Paap F-statistics based on Conley standard errors are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Correlation between the instrument and the sum of shares

	Instrument		Average Immigrant Share, 1860-1920		Log income per capita in 2000	
	(1)	(2)	(3)	(4)	(5)	(6)
Expected instrument (μ_i)	0.954***	1.332***	4.481***	4.925***	11.001***	20.857***
	(0.014)	(0.071)	(1.167)	(2.136)	(2.364)	(7.898)
Observations	2935	2935	2935	2935	2935	2935
Baseline controls		✓		✓		✓

Notes: OLS regressions of the instrument, average migration share between 1860 and 1920 and log income per capita in 2000 on the sum of shares. Odd columns do not include any controls; even columns include all baseline controls of Table 4 of SNQ, including the *log(years of rail+1)* control. Conley standard errors using a five-degree window are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Controlling for μ_i and additional controls

	(1)	(2)	(3)	(4)
Panel A: First stage	Average immigrant share, 1860–1920			
Instrument	4.610***	5.653***	3.707**	4.798**
	(1.465)	(1.923)	(1.460)	(1.977)
Panel B: Reduced Form	Log income per capita, 2000			
Instrument	6.560*	3.243	3.669	-1.627
	(3.790)	(5.585)	(3.219)	(5.303)
Panel C: 2SLS	Log income per capita, 2000			
Average Immigrant	1.423*	0.574	0.990	-0.339
	(0.875)	(1.016)	(0.950)	(1.096)
RI p-value	{0.821}	{0.847}	{0.931}	{0.933}
Anderson-Rubin 90% CI	[0.09, 3.54]	[-1.14, 2.91]	[-0.45,3.84]	[-2.76,2.1]
Observations	2935	2935	2935	2935
First stage F	9.91	8.64	6.45	5.89
R^2 of IV on controls	0.945	0.972	0.942	0.973
Years of railroad connection	✓	✓		
Years since county establishment			✓	✓
Sum of shares μ_i		✓		✓

Notes: First stage, reduced form and 2SLS estimates with alternative controls for timing of railroad connection and county age. All models include the baseline controls of Table 4 of SNQ, except for the *log(years of rail+1)* control. Columns 1 and 2 controls linearly for the number of years since railroad connection and columns 3 and 4 for the number of years since county establishment. Columns 1 and 3 additionally control for the sum of shares. Conley standard errors using a five-degree window are shown in parentheses. Anderson-Rubin confidence intervals (square brackets) and Randomization Inference p-values (curly brackets) are reported for the key 2SLS coefficient of interest. Kleibergen-Paap F-statistics based on Conley standard errors are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.