The Train Wrecks of Modernization: Railway Construction and Separatist Mobilization in Europe

Yannick I. Pengl *ETH Zurich*

yannick.pengl@icr.gess.ethz.ch

Carl Müller-Crepon *LSE*

c.a.muller-crepon@lse.ac.uk

Roberto Valli *ETH Zurich*

roberto.valli@icr.gess.ethz.ch

Lars-Erik Cederman *ETH Zurich*

lcederman@ethz.ch

Luc Girardin *ETH Zurich*

luc.girardin@icr.gess.ethz.ch

September 29, 2023

Abstract. This paper uses the gradual expansion of the European railway network 1816–1945 to investigate how this key technological driver of modernization affected ethnic separatism. Combining new historical data on ethnic settlement areas, conflict, and railway construction, we test how railroads affected separatist conflict and successful secession as well as independence claims among peripheral ethnic groups. Difference-in-differences, event study, and instrumental variable models show that, on average, railway-based modernization increased separatist mobilization and secession. Exploring causal mechanisms, we show how railway networks can either facilitate mobilization by increasing the internal connectivity of ethnic regions or hamper it by boosting state reach. As expected, separatist responses to railway access concentrate in countries with small core groups, weak state capacity, and low levels of economic development as well as in large ethnic minority regions. Overall, our findings call for a more nuanced understanding of the effects of European modernization on nation building.

. Previous versions were presented at EPSA in June 2023; Oxford University on March 20, 2023; the London School of Economics on March 21, 2023; University of Gothenburg, March 2023; University College London

Nineteenth and early twentieth century Europe saw unprecedented economic, political, and cultural change. Industrializing economies, expanding markets, centralizing states, and nationalist ideologies fundamentally transformed both private and public life (Osterhammel 2014; Buzan and Lawson 2017; Ansell and Lindvall 2021). New transport technologies, especially railways, were one key lever that drove and connected these modernizing forces (Maier 2016). Railroads connected previously isolated subnational regions, fostered industrialization, and boosted the state's ability to reach and govern peripheral populations. As such, they helped to create the communicative, economic, and political conditions that, according to classic modernization theories, promoted national integration and identity formation (Deutsch 1953; Gellner 1983; Anderson 1983). At the same time, expanding transportation networks contributed to the mobilization of culturally distinct peripheral groups, spurring instability and local backlash against the economic and political transformations they brought about (Hechter 2000; Breuilly 1982; Huntington 1968).

In this paper, we investigate how the expanding European railway network contributed to nationalist mobilization and whether it united or divided existing states. Our theoretical argument builds on and extends the existing literature on modernization, nationalism, and separatism. We specify three specific mechanisms through which railroads may affect competition and bargaining between the central state and an ethnically distinct peripheral region. While improved access to national markets and the capital city can be expected to promote integration and stability, internal connections in the periphery are likely to fuel local mobilization and

Conflict & Change group in January 2023; the Hertie School in Berlin, November 2022; the Graduate Institute in Geneva, October 2022; APSA in Montreal, September 2022; at the Princeton LISD and IPERG-UB Workshop on National Identity in Barcelona, May 2022; and at the Zurich Political Economy Seminar Series, June 2022. We thank the discussants (Idil Yildiz, Scott Gates, Patrick Kuhn, Leonid Peisakhin, Bruno Caprettini) and participants of these workshops for their valuable comments and suggestions. This research is supported by the Advanced ERC Grant 787478 NASTAC, "Nationalist State Transformation and Conflict."

separatism. Since the integrative processes of cultural assimilation, state-led nation-building, and economic modernization tend to unfold over longer time frames than local resistance, we expect the first arrival of rails in ethnic minority regions to increase the risk of separatist mobilization. The impact of more gradual extensions of the network is likely to depend on how they affect national market access and state reach on the one hand, and local mobilization capacity on the other. In addition, we derive hypotheses on how ethnic demography, economic development, and political institutions may make a difference in whether railroad construction led to national integration or disintegration.

We test these arguments by combining newly collected geo-spatial data on the expanding European railway network with information on separatist mobilization in the form of national independence claims, secessionist civil wars, and successful secession. We link these data to yearly observations of ethnolinguistic group segments which we derive by intersecting detailed historical maps of ethnic settlement patterns with time-varying country borders covering the period 1816–1945.

First, we find that, on average, railway access is associated with an about twofold increase in the probability of separatist mobilization. Consistent with out theoretical reasoning, this effect materializes immediately and, if anything, dissipates over time although it never turns negative. Second, our analysis reveals that improvements in state reach reduce separatism, internal connectivity increases the risk, and market access exerts no significant effects. In a third step, we show that separatist responses to railway access concentrate in contexts that complicate top-down nation-building and provide motivations and opportunities for national independence campaigns. Finally, since railway construction could be in itself motivated by the threat of separatism, we also account for reverse causation by creating a simulated instrument based on the assumption that the rail network serves to connect a country's population as efficiently as possible. The results of the IV analyses are well in line with our baseline findings and make it unlikely that we merely pick up strategic railroad building in areas at high risk of conflict.

Our paper contributes to the literatures on modernization, nationalism, separatism, and the political consequences of transport and communication technologies. Analyzing railroad construction and other dimensions of modernization, historians provide convincing qualitative evidence on national integration in France on the one hand (Weber 1976) and disintegration and separatist nationalism in Eastern Europe on the other (Breuilly 1982; Connelly 2020). In economic history and geography, there is a rich literature on the impact of railway construction on economic development, urbanization, and industrialization (see, e.g. Fishlow 1965; Hornung 2015; Berger 2019; Alvarez-Palau, Díez-Minguela, and Martí-Henneberg 2021; Donaldson and Hornbeck 2016; Donaldson 2018), but much less is known about how it influences political outcomes, such as nation-building. In a study of 19th century Sweden, Cermeño, Enflo, and Lindvall (2022) show how railways empower public school inspections, leading to higher enrollment rates and more nationalist curricula in connected locations. On the flip side, recent empirical contributions link railroads to the diffusion of opposition movements in specific European and non-European countries (Brooke and Ketchley 2018; García-Jimeno, Iglesias, and Yildirim 2022; Melander 2021).[1]

What is missing, however, are studies that analyze both integrative and disintegrative dynamics systematically and more broadly across Europe. Our arguments and findings provide a systematic and comprehensive assessment of how a crucial technological driver of modernization relates to the rise and spread of separatist mobilization across historical Europe.

Modernization and Nationalism in the literature

In this section, we briefly situate our paper within the broader literature on modernization and nationalism from which we will draw in the theory section below. The introduction of steam-

1. For studies on more recent communication technologies and their impacts on national identification, political mobilization, and conflict, see e.g. Choi, Laughlin, and Schultz (2021), Pierskalla and Hollenbach (2013), Shapiro and Weidmann (2015), Christensen and Garfias (2018), Enikolopov, Makarin, and Petrova (2020), Gohdes (2020), and Manacorda and Tesei (2020).

powered railroads is often described as "the defining innovation of the First Industrial Revolution" (Cermeño, Enflo, and Lindvall 2022, 715) and is thus inextricably linked with the various modernization processes that spread across Europe in the 19th and early 20th centuries. A large, and by now classic, literature links the rise of nationalism to these processes (see e.g. Deutsch 1953; Gellner 1983; Anderson 1983). At the risk of some simplification, the relevant arguments can be organized into two main camps depending on whether they stress national integration or separatism as main effect of modernization. This dimension relates to the prominent distinction between "state-building" and "peripheral" (Hechter 2000) or "state-framed" and "counter-state nationalism" (Brubaker 1996).

The classic contributions are typically interpreted as predicting cultural homogenization and increasing identification with the state-leading nation (see e.g. Robinson 2014; Eifert, Miguel, and Posner 2010). Explicitly political accounts stress the modern state as the key agent of change and top-down social engineering (Hobsbawm 1990). On this view, states and their rulers devise and implement far-reaching nation-building programs to respond to both international and domestic threats to their political survival (Hintze 1975; Tilly 1994; Posen 1993; Alesina, Giuliano, and Reich 2021). A complementary, equally prominent perspective views the development of industrial economies rather than instrumentalist state policy as the main cause of national integration. In Gellner's (1983) seminal account, the transition from agrarian to industrial modes of production requires literate workers and professionals speaking and writing in standardized languages in large and integrated national economies (see also Gellner 1964; Green 2022). In a pioneering book, Deutsch (1953) highlights expanding communication networks resulting from technological innovation, labor migration, and market exchange as major industrial driver of nationalism. Similarly, Anderson (1983) describes how "print capitalism," that is the mass production and consumption of newspapers, novels, and educational materials in standardized vernaculars, leads ordinary people to imagine themselves as members of broader, increasingly national communities.

Contrary to what simplistic interpretations would suggest, modernist accounts also provide useful starting points to understand national disintegration. Adopting a largely state-centric perspective, Breuilly (1982) and Hechter (2000) expect the shift from indirect to direct rule to trigger reactive mobilization, especially in situations where peripheral elites enjoyed considerable autonomy prior to state centralization (see also the notion of "internal colonialism," Hechter 1977). Relatedly, Deutsch (1953) notes that wherever communication-induced social mobilization outpaced assimilation into dominant national cultures, local backlash and nationalist conflict became more likely. Gellner (1983, 1964) expects the combination of pre-existing cultural difference and uneven development to trigger separatism which he refers to as "Habsburg nationalism."

Beyond the theoretical classics, several empirical studies shed valuable, but more selective, light on the link between modernization and nationalist mobilization. Perhaps most famously, Weber (1976) carefully traces French national identity formation in the late 19th century highlighting industrialization, extended transportation and communication networks, as well as instrumental state policy as integrating forces. Despite the brilliance of his contribution, Weber (1976) remains a historian of France, a country that is quite the outlier in terms of successful nation-building, even when judged by stringent Western European standards.

More recently, cross-country studies have shown that state-led nation-building efforts, in particular education reforms, become more likely when rulers face international (Darden and Mylonas 2016; Aghion et al. 2019) or domestic threats (Paglayan 2022; Alesina, Giuliano, and Reich 2021).[2] While these studies offer information about the strategic timing of nation-building policies, the mere adoption of such efforts does not guarantee their success. They may lead to successful assimilation or backlash and the outcome likely depends on contextual variation between and within nation-building states.

2. But see Garfias and Sellars 2021a who find that in colonial Mexico, the transition to direct rule occurred faster in subnational areas with lower risk of internal resistance.

Micro-level quantitative work within single countries illustrates how specific educational, linguistic, religious, and state-building efforts succeeded or backfired in 19th century and contemporary France (Balcells 2013; Blanc and Kubo 2022; Abdelgadir and Fouka 2020), Prussia (Cinnirella and Schueler 2018), colonial Mexico (Garfias and Sellars 2021b), the early 20th century US (Fouka 2020), and Atatürk's Turkey (Assouad 2020). These contributions provide important evidence on how specific state policies cause national integration or backlash but say less about variation across contexts.

To our knowledge, the only comparative study that offers broad geographic scope is the aforementioned article by Wimmer and Feinstein (2010). Their analysis focuses on nation-state creation in a global sample of 145 territories corresponding to independent states in 2001 back-projected until 1816. Relying on railway density as a modernization proxy in pooled time-series regressions, they fail to detect any effects on the transition to nation-states in existing pre-national or newly independent states. While making a pioneering effort to systematize empirical analysis, their country-level approach remains both over-aggregated and merely correlational. Most seriously, their research design suffers from hindsight bias due to the backward-projected sampling based on contemporary state units.

In sum, then, the link between modernization and national integration and disintegration remains contested, and lacks contextualization and empirical evidence that bridge the micro- and macro-levels. First, there is significant disagreement as to whether modernization spurs nationalism in support of, or against, the state and which political or socio-economic mechanisms account for any association between modernization processes and different forms of nationalist mobilization. Second, and as an important potential source of this disagreement, the existing literature provides little theoretical guidance or empirical evidence as regards the contextual factors that let state-building or counter-state nationalism prevail in specific cases.[3] Third, the

3. Hechter (2000, 93), for example, expects peripheral nationalism where "the institutions of direct rule fail to assimilate the distinct nations within a state." Where, when, and why assimilation fails, however, remains to be theorized.

classic contributions and more recent micro-level studies leave an important empirical middle ground uncovered. While the former do not provide systematic evidence for their claims,the latter convincingly validate parts of the classical theories in selected European countries but lack a broader comparative outlook.

The present paper addresses these three gaps in the existing literature. First, we zoom in on railway construction to assess whether this crucial technological driver of political and economic modernization in 19th and early 20th century Europe is systematically related to processes of national integration or disintegration. Second, we derive and test hypotheses on causal mechanisms as well as country-level and subnational contextual factors that make national integration or counter-state nationalism more likely. Third, our treatment and outcome data are spatially disaggregated at the subnational level and cover the entirety of Europe. These data allow us to integrate the two dominant types of analysis in the existing empirical literature, represented by cross-country comparisons and micro-level work on individual cases.

Theory: Railways and nationalist mobilization

As the short discussion of existing research makes clear, railway expansion and the associated local modernization processes likely affected European nationalisms through multiple mechanisms and with ambiguous implications for national cohesion and political stability within given state borders. The integrative potential of expanding state presence as well as the exchange of goods, people, and ideas over large distances point to successful nation building. At the same time, local connectivity and modernization may facilitate mobilization of opposition and spur separatist responses to state expansion and national integration.

Why separatism?

Our theoretical framework draws on this literature to explain how and under which conditions railroad construction united or divided historical Europe's at least initially multi-ethnic states. More specifically, we lay out mechanisms through which railways affect the motivations and opportunities for separatist mobilization among non-core population groups. These groups are culturally distinct from their host state's governing elites and typically smaller in demographic size as well as more peripherally located than their state-leading counterparts (Mylonas 2012). With the exceptions of small and culturally homogeneous polities in pre-unification Italy and Germany, practically all states in historical Europe contained notable non-core population segments. In pre-industrial times, central state governments typically ruled non-core groups in indirect fashion outsourcing important governing tasks to local intermediaries (Hechter 2000). Cultural difference and mediated forms of projecting power suggest that most European states still operated more like empires (Motyl 1997; Burbank and Cooper 2010).[4]

The situation changed when industrialization, increasingly direct forms of rule, and nationalist ideologies swept across Europe in the 19th century. Separatist mobilization became more frequent and occurred wherever non-core group leaders managed to rally a sufficient number of followers around movements that called for national independence. Such leaders were, on the one hand, old elites who had benefited from agrarian economies or indirect rule arrangements and whose status was threatened by local industrialization or state centralization (Hechter 2000; Garfias and Sellars 2021b). On the other hand, there were new elites ranging from bourgeois liberals and democratic reformers to cultural awakeners and ethnonationalists who frequently came to see independence as a route to fulfill both their personal aspirations and their political ideas (Gellner 1983).

4. Historians refer to these units as "composite monarchies" (Elliott 1992). Even metropolitan France, arguably the most advanced case in terms of state centralization and national sentiment in the early 19th century is prominently described as an essentially imperial setting (Weber 1976).

For old and new non-core elites, separatism provided several advantages over alternative forms of mobilization. First, national independence would assure exclusive access to the benefits of local governance which were increasingly endangered by central state expansion (Hechter 2000). Second, stressing cultural unity at the local or regional level helped to forge coalitions between old agrarian elites and rising middle classes whose economic interests were typically not aligned (Breuilly 1982). Third, and as soon as ideologies of popular sovereignty and national self-rule took root, bravely resisting domination by a culturally foreign host state elite allowed them to mobilize local populations more effectively than alternative opposition frames (Balcells, Daniels, and Kuo, n.d.; Gellner 1983). In addition, separatist mobilization raised the prospects of securing support from nationalizing Europe's great powers, which became increasingly receptive to national self-determination as legitimate cause for political action (Breuilly 1982).

Taking separatism as the main outcome under investigation circumvents the challenge of adequately defining and measuring national integration at subnational levels. National integration can be achieved through assimilation into the national core group, the development of an overarching identity on top of ethnic diversity, or political integration and power sharing across ethnic divides (Wimmer 2018; Rohner and Zhuravskaya 2023). Given these different paths to national cohesion, it seems analytically more productive to focus on clear failures of nation building. The downside of this approach is that the reverse, however, is not true, as the absence of separatist mobilization does not imply that national integration has been achieved (Connor 1972). We thus have to make the untested assumption that railroads can have integrative effects and that, all else equal, these effects should reduce separatism among non-core populations.

Motivations driving railroads construction

Before discussing the consequences of railroads in Europe, we provide a brief overview of the motivations behind their construction. In Britain, commercial actors took the pioneering steps toward connecting urban centers (Trew 2020; Bogart 2009). The British case, however, is rather

an outlier in this respect. The French trajectory is more representative of the the political logic that dominated in continental Europe. As shown by Eugen Weber (1976), France saw a more active governmental role in railway planning, which served to promote not only economic development but also national integration and cultural penetration into the periphery of the state's territory. The centralizing logic was also present in Sweden (Cermeño, Enflo, and Lindvall 2022), Belgium, and with considerable delay Spain (Alvarez-Palau, Díez-Minguela, and Martí-Henneberg 2021). In unifying Germany and Italy, railroad construction was enlisted in the quest of integrating previously independent entities, although with considerable lack of efficiency in the latter case (Schram 1997). French planners were also motivated by geo-strategic considerations, especially the need to counter Prussian/German rail-based mobilization (e.g., Alvarez-Palau, Díez-Minguela, and Martí-Henneberg 2021, 264).[5]

Further east, the large multi-ethnic empires were more reluctant to engage in nation-building. Their dynastic elites saw nationalism primarily as a threat rather than as an asset. Together with lagging development and an uneven access to capital, this reluctance delayed the introduction of railways and their use for the purpose of nation-building. Nonetheless, the military threat posed by the western great powers increased the pressure on imperial decision making, both in the Habsburg Empire and tsarist Russia (Gutkas and Bruckmüller 1989). While commercial interests had driven early railroad construction in the former empire, concerns with securing its borders and quickly deploying its troops motivated Vienna's extension of railroad lines to the Russian border and into the Italian peninsula (Köster 1999; Rieber 2014).

With even less access to private finance, the Romanov Empire similarly used railways to reinforce its external borders, but equally prominently as a tool of imperial rule (Schenk 2011). In 1863, the newly built rail connection between St. Petersburg and Warsaw allowed the tsarist regime to send troops that crushed the Polish revolt. At the same time, the belated drive for nation-building and Russification gave railroads a prominent role as cultural homogenizers.

5. Yet, as such, railway construction fueled by interstate warfare is unlikely to drive reverse causation since this study focuses on internal and secessionist conflict.

Foreboding Putin's more recent expression of imperialist attitudes, Baron Korf, who visited Ukraine after the Polish rebellion, predicted a harmonious path toward national integration:

> The way to it is through the railroad. . . . It is not only the goods that move along this road, but also books, ideas, customs, views. . . . The Great Russian and Little Russian capitals, ideas, views will mix, and these two peoples, already standing close, will first become related and then become one (quoted in Miller 2015, 359–360).

Below, we will return to the reasons motivating railroad construction in our discussion of endogeneity.

Railroads, modernization, and separatist mobilization

We now turn to our main arguments of how railroad construction may affect the choice of non-core populations to join and support separatist movements. This choice depends on the expected costs, benefits, and chances for success of state-led nation-building on the one hand and national independence campaigns on the other. Railway construction in the periphery may thus affect the emergence of separatist movements if it significantly shifts these costs, benefits, and success probabilities as perceived by local populations. In what follows, we describe three broad mechanisms through which access to expanding railway networks matters and derive our baseline hypothesis. Next, we link our causal mechanisms to specific forms of more gradual railway expansion, and, in a last step, derive contextual factors that may tilt the balance in favor of the central state or separatist movements.

The three main theoretical mechanisms through which railroads may have affected non-core individuals in modernizing Europe are illustrated in Figure 1 and relate, respectively, to increased interactions between core and non-core groups (M1), the state's ability to reach and penetrate non-core populations (M2), and non-core elites' and populations' capacity to mobilize

against the state (M3). M3 can be expected to unambiguously increase the risk of separatism. M1 and M2 are likely too reduce separatist mobilization in the long term, but may involve short-term backlashes and instability.

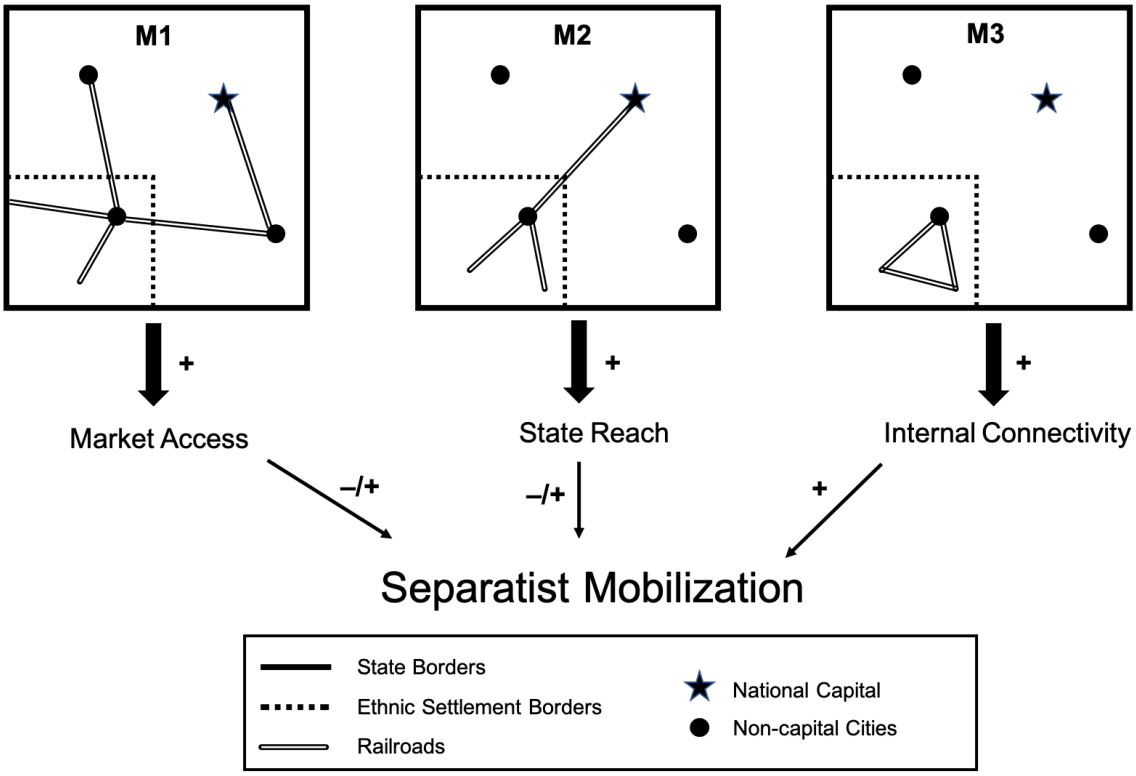


Figure 1: How railroad construction may matter

M1: Market access and social communication. First and foremost, railroads affect local populations through economic integration and social communication. Improved connectivity throughout a country's territory, and especially to major cities, provides peripheral populations with material incentives to orient themselves toward an increasingly national economy and, in some cases, to even culturally assimilate into supralocal national identities. Mechanism M1 in

Figure 1 schematically illustrates this point. The two railroad lines directly link the non-core population segment in the bottom-left corner to the two non-capital cities.

Industrial and capitalist development is inextricably linked with railway construction. Moving goods and people across distance enabled the formation of integrated market economies and labor migration from agrarian towns to industrializing cities (Rostow 1960; Fishlow 1965; Weber 1976). Recent empirical studies show how railway building contributed to city growth, increasing employment shares in the industrial sector, and more integrated markets in 19th century Europe (Keller and Shiue 2008; Hornung 2015; Alvarez-Palau, Díez-Minguela, and Martí-Henneberg 2021; Berger and Enflo 2017; Berger 2019). By the same token, urbanization and industrialization spurred railway construction as the earliest lines typically connected the major industrializing cities within a country (Hornung 2015). Where railways brought income earning potential and prospects for upward mobility within national markets, local residents were unlikely to support peripheral elites' attempts to cut them off from these emerging opportunities by breaking away from the state (Hierro and Queralt 2021).

Railways accelerated the expansion of communication networks, brought previously isolated rural residents in contact with urban dwellers and each other, and can thus be expected to have created the bottom-up incentives and pressures for cultural homogeneization described by Gellner (1983) and Deutsch (1953). Not surprisingly, Weber (1976, ch. 12) describes road and railway networks as technological precondition for "radical cultural change" in nationalizing France Segal (see also 2016). Maier (2016) even uses the term "railroad nationalism" to describe the transformative effects of the transport revolution on national integration in Europe and the United States.

However, cultural difference may become more salient where members of distinct ethnic groups compete for inherently scarce modernization benefits (Bates 1983). Similarly, Gellner (1983) explains how economic integration and information flows can make ethnically distinct peripheries acutely aware of their subordinate status and limited prospects for upward mobility

which should increase support for separatist movements. Railroad expansion could thus also be expected to increase peripheral populations' motivations and elites' opportunities for separatist mobilization.

M2: State reach and direct rule. A second and plausibly equally important mechanism links railroads to the central state's opportunities to reach, govern, and transform local populations in top-down fashion. Providing public goods and engaging in ambitious nation-building policies would have been inconceivable without railroads (Wimmer 2018). Modern transportation infrastructure significantly boost what Mann (1993, 59) calls the "infrastructural power" of European states, which he defines as the "institutional capacity of a central state [...] to penetrate its territories and logistically implement decisions." Mechanism M2 in Figure 1 depicts this logic with a direct railroad link from the national capital to the main city in the culturally distinct non-core region. Here again, both local and non-local railway building matters as each kilometer of tracks constructed between the capital and the non-core segment implies reduced travel times from the political center.

Central states need to reach and penetrate peripheral areas to implement their preferred policies, monitor state-appointed bureaucrats, and, if necessary, repress unruly local elites and populations (Hechter 2000, 29). Cermeño, Enflo, and Lindvall's (2022) analysis of 19th century Sweden supports this view, showing how railways enabled public school inspectors to better reach peripheral districts, leading to higher enrollment rates and more nationalist curricula in connected locations. If railway-enabled public goods provision (Wimmer 2018; Alesina and Reich 2015), mass education (Paglayan 2021, 2022; Alesina, Giuliano, and Reich 2021), and policing capabilities (Mann 1993; Müller-Crepon, Hunziker, and Cederman 2021) induce loyalty and obedience as intended, local populations should have reduced motives and opportunities to support separatism.

At the same time, however, increasing state penetration and top-down nation-building (M2 in Figure 1) may spur backlashes where they proceed—or are perceived—as exploitative schemes of "internal colonialism" (Hechter 1977). In addition, the mere fact of "alien rule" by ethnically distinct central state elites, regardless of specific policies, appeared increasingly scandalous in nationalizing Europe (Hechter 2013). By bringing the state closer to peripheral elites and populations and thus threatening their status, power, and traditional ways of life, railroad networks can plausibly contribute to the emergence of "reactive nationalism" (Hechter 2000).

M3: Internal connectivity and social mobilization. Third, railroads may also facilitate the coordination and collective action of peripheral opposition movements. Mechanism M3 in Figure 1 shows how local rails within the culturally distinct subregion of the country improve the internal connectedness of its residents. Rapidly spreading information and ideas as well as social ties between leaders, activists, and ordinary citizens are key ingredients to successful mobilization (Granovetter 1978; Kuran 1992; Shesterinina 2016; Aidt, Leon-Ablan, and Satchell 2022). In line with this notion, recent empirical studies illustrate how railroad connectivity contributed to the diffusion and growth of opposition movements in the 19th-century United States (García-Jimeno, Iglesias, and Yildirim 2022), pre-democratic Sweden (Melander 2021) and interwar Egypt (Brooke and Ketchley 2018). Specifically related to nation building, Deutsch (1953) expects ethnic conflict where social mobilization through improved communication happens before local assimilation into dominant national cultures. By boosting internal connectivity, local railroad construction may thus increase the opportunities for separatist mobilization and, via internal communications and exchange, promote identification with separatist movements.

Deriving testable hypotheses. Taken together, the three causal mechanisms just outlined generate ambiguous expectations as regards the link between railroad construction and separatism. On the one hand, railways provide the transportation and communication networks that integrationist modernization theories regard as essential for both bottom-up (M1) and top-

down nation building (M2). On the other, both market integration and state penetration may spur local backlashes and internal connections (M3) are likely to facilitate separatist mobilization. There are, however, several reasons to expect railroad construction in non-core areas to increase the risk of separatism, at least in the short term.

First, and as illustrated in Figure 1, newly built rails within the settlement area of a non-core group unambiguously improve internal connectivity, whereas market access and state reach also depend on non-local railways in other parts of the country. Second, both the market access and the the state reach mechanism do not unequivocally point to integration but may also foster resistance and separatist mobilization. Third, the integrative and assimilationist effects of market integration, social communication, and state reach typically unfold gradually and only fully materialize in the longer term. Economic change and local industrialization tend to uproot local modes of production and systems of exchange before adaptation is complete and the benefits trickle down to broader segments of the local population. While contact and exchange through personal mobility and labor migration have the potential to foster cultural homogenization into overarching national identities, such cultural change rarely occurs overnight (Weber 1976). Similarly, state-led nation-building policies such as mass schooling and compulsory military service target younger generations and will therefore take full effect decades after their first introduction (Blanc and Kubo 2022).

Thus, we expect the first railway connections in non-core regions to increase the risk of separatism. The effects of internal connectivity on coordination and social mobilization likely materialize in more immediate fashion than the integrative forces described above. In addition, where local elites and populations regard incipient economic change and state penetration as threats, they face strong incentives to mobilize resistance before slow-moving assimilationist

pressures undermine their local basis of support. We therefore state our first hypothesis as follows:

Hypothesis 1. Railway construction in non-core regions increases the likelihood of separatist mobilization, at least in the short term.

The first task of the empirical analysis below is thus to test if there is any systematic relationship between local railroad construction and peripheral nationalism and, if yes, whether a first railway connection increases the potential for counter-state nationalism as hypothesized. To leave it at that, however, would be theoretically unsatisfying. European history provides numerous examples of both successful nation building and national disintegration. The conditions under which one or the other prevails appear as an equally, if not more, important puzzle than any general relationship between railroads and separatism.

Network structure and specific causal mechanisms. To make progress toward resolving this puzzle, we move beyond the mere presence of a railway connection and investigate how more gradual and long-term improvements in connectivity relate to three mechanisms described above. The main drivers in bottom-up versions of integrationist modernization theory are industrial development, urbanization, as well as personal mobility and exchange over larger distances. This mechanism (M1 in Figure 1) should be particularly relevant where railway construction effectively integrates peripheral regions into national markets and improves local population's access to the industrializing cities of the country. Provided that they do not trigger inter-group conflict or competition, railway lines that increase a region's "market access" (Donaldson and Hornbeck 2016) can be expected to lower local incentives for separatism and contribute to growing identification with the state-framed national identity, especially in the long run.

In similar vein, top-down nation building through public goods provision, education, and repression requires fast and reliable transportation links between the state capital and poten-

tially restive minority regions (M2). Separatist mobilization therefore seems less likely wherever newly constructed rails more directly connect peripheries with the administrative capital and the integrative effects of direct rule and top-down nation-building prevail over local efforts to mobilize for separatism (M2 in Figure 1).

In addition, new transportation links can also boost internal connectivity within peripheral regions without simultaneously increasing state reach or national market access (M3 in Figure 1). We thus test the following three, more long-term hypotheses linking the structure of expanding European railway networks to the likelihood of separatist mobilization.

Hypothesis 2 Railway-induced improvements in ...

(a) ...national market access reduce the likelihood of separatist mobilization (M1).

(b) ...state reach reduce the likelihood of separatist mobilization (M2).

(c) ...internal connectivity increase the likelihood of separatist mobilization (M3).

Conditional hypotheses. Beyond network structure and different types of connectivity improvements, specific contextual conditions are likely to shape the opportunities and motivations for separatist mobilization. We explore five cultural, demographic, political, and economic factors that either complicate top-down nation building or favor separatist mobilization.

First, large cultural distances make it harder for the state to reach, govern, and assimilate peripheral populations (Alesina and Reich 2015). Homogenizing populations speaking local dialects of the dominant language or at least belonging to the same linguistic family appears easier than bridging deeper cultural divides

Second, where large majorities already speak some version of the state-sanctioned national language, the standardization across local dialects and assimilation of culturally more distinct but small national minorities becomes a realistic prospect. Conversely, national integration

appears a much more daunting task where the state-leading nation represents relatively small shares of its country's population.

Third, national independence campaigns only gain support where they can mount a credible challenge to the host state and offer the prospect of economic and military viability in case of successful secession (Siroky, Mueller, and Hechter 2016). Non-core groups with large populations and territories can more credibly promise sufficient state and market size after independence and are therefore more likely to rally the required support than small national minorities (Hechter 2000, ch. 5).

Fourth, in underdeveloped countries, railway access likely brings in the central state but does not come with the economic benefits and opportunities of rapid industrialization, peripheral populations have little incentives to become loyal to the center or invest in cultural assimilation. Under such conditions, claims about exploitation by the ruling elite are particularly likely to resonate with local populations (Hobsbawm 1990, ch. 4).

Fifth, only high-capacity states can be expected to successfully implement direct rule and ambitious nation building policies. Pre-existing levels of state and especially fiscal capacity developed through earlier processes of political reform, technology adoption, or economic integration are thus likely to matter (Wimmer 2018).

Last but not least, democratic institutions, especially liberal ones that protect all and, in particular, minority citizens against excesses of the state might make peripheral populations more likely to accept or even support direct rule by the center.

Based on these contextual arguments, we specify and test additional hypotheses on the link between railroads and separatism.

Hypothesis 3. Railway access increases the likelihood of separatist mobilization in

(a) non-core groups that are culturally distant from the state-leading nation,

(b) countries dominated by a relatively small national core group,

(c) large non-core groups,

(d) relatively poor and less industrialized countries,

(e) low-capacity states,

(f) staunchly autocratic states.

Data and variables

Our analysis requires a geographic unit of analysis below the country level from which separatist mobilization against the state likely emanates. In all analyses, we use yearly observations of ethnic segments, defined as the spatial intersections between country borders and ethnic settlement areas.

Ethnic settlement data.

Information on historical ethnic settlements comes from the newly compiled Historical Ethnic Geography (HEG) dataset. This dataset is based on a selection of 73 historical maps (for details, see Appendix). Practically all ethnic categories appearing on our maps refer to linguistic rather than religious or regional ethnic identity markers. That said, some maps differ in the level of linguistic granularity they encode and therefore need to be standardized. To address this "grouping problem'' of European ethnolinguistic identities, we match all raw linguistic map

labels to the Ethnologue language tree (Lewis 2009) and construct a time-invariant master list of relevant groups. Finally, we draw on all maps belonging to a specific group-time period combination to construct a best-guess settlement polygon.

Any data on ethnic settlements covering as broad a geographic and temporal scope as 19th and 20th century Europe are prone to some measurement error. We address this challenge by pre-selecting only the highest quality maps, hand-coding periods of significant change, and combining information from multiple maps. These steps ensure a relatively accurate dataset and minimize concerns about systematic biases in our units of analysis.

Other concerns relate to endogenous ethnic settlement areas and sample selection. Ethnic geography may be affected by past conflict, nation-building policies, and other political forces. Our historical maps might also miss small and extinct groups that were assimilated into broader national or linguistic categories. As a result, large and politically mobilized ethnicities are likely to be overrepresented in our sample. Since these groups are bigger and more likely to be active in politics, they can be expected to have a higher baseline risk of making secessionist claims or being involved in territorial conflict. If relevant, this selection issue should make it harder to identify effects on conflict and separatism.

Historical state borders.

Spatial data on state borders since 1886 come from the CShapes 2.0 dataset that offers global coverage on all sovereign states and their dependencies since the "Scramble for Africa" (Schvitz et al. 2022). These data were extended for Europe back to 1816 drawing on non-spatial data from the Gleditsch and Ward (1999) dataset of independent states, the Correlates of War's Territorial Change dataset (Tir et al. 1998), and historical GIS data from the Centennia Historical Atlas (Reed 2008), with the addition of dozens of microstates that existed before the German and Italian unifications.

Units of analysis

Spatially intersecting the aggregate group polygons with yearly data on European state borders yields our main unit of analysis – ethnic segments years (*ect*) starting in 1816. For each segment year, we calculate absolute area and population. Historical population data comes from the History Database of the Global Environment (HYDE) (Goldewijk, Beusen, and Janssen 2010). Wherever ethnic segment or aggregate group polygons overlap, we equally divide area or population between overlapping polygons.

As national core groups do not engage in separatism, our baseline analyses restrict the sample to non-core ethnic segment years. Core groups are identified as the ethnic segment with most power in the respective country's capital. The largest ethnic segment that contains the capital serves as our first guess which we then manually correct wherever necessary.

Main independent variable: Railway access

We use local-level access to railway networks as geographically and temporally disaggregated proxy for the uneven spread of modernization across Europe. Geographic data on the expanding European railway network comes from train.eryx.net, an antiquated website built by French train enthusiasts Bernard and Raymond Cima. They provide construction dates and map representations of all known railway segments covering almost the entirety of geographic Europe, with the notable exception of England and Wales, which we exclude from the analysis. We georeferenced their yearly online map tiles and digitized all line features to construct a geospatial dataset of European rails covering the period 1834 to 1945. Figure 2 plots our railroad data. In the Appendix (A2) we validate the railroad data's precision against a set of time-varying railway maps for Austria-Hungary.

The main treatment indicator in the analyses below is a dichotomous railroad access indicator derived from intersecting the yearly ethnic segment polygons with yearly spatial line datasets

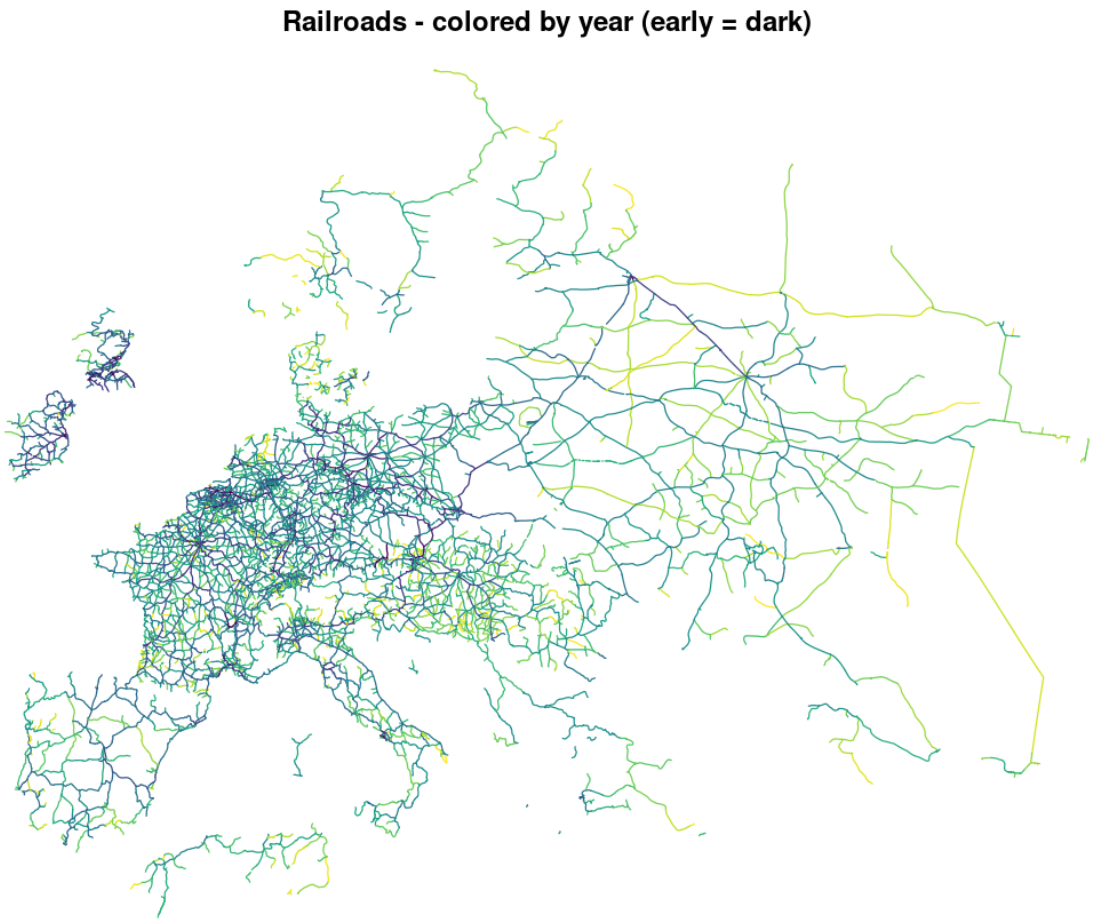


Figure 2: Geographic data on yearly railway construction (digitized from train.eryx.net).

of the European railway network. All segments intersected by a line feature are assumed to be connected. To get closer at causal mechanisms, we use the network structure of the railway data to compute continuous proxies for state reach and connectivity to urban markets (see Appendix A3 for details).

Outcomes: Attempted and/or successful secession

As described below, our main outcome variable captures violent and peaceful mobilization for separatism by combining successful secessions, onsets of separatist conflict, and political claims for national independence or regional autonomy.

Secessionist Conflict. First, we code a dummy of territorial ethnic civil war onsets at the ethnic segment-year level. For the period 1816–1945, we identify all unique civil wars listed in the datasets provided by Gleditsch (2004) and Sarkees and Wayman (2010) that are fought in the name of a specific ethnic group. The coding rules for linking conflicts to ethnic segments are the same as in the ACD2EPR dataset, requiring explicit ethnic claims and recruitment from a particular ethnic group (Wucherpfennig et al. 2012). We restrict the main analysis to secessionist conflict.

Successful secession. We combine the secessionist conflict measure with a binary indicator of successful secession as an additional signal of national disintegration.[6] The secession dummy is coded one for all non-core ethnic segments that become core group segments in newly independent states in year $t+1$. In combination, the disintegration measure takes on the value of 1 if a segment experiences a secessionist conflict onset or secedes in a given year and 0 otherwise.

National independence and autonomy claims. As a complement, we also use new dataset to code the first claim for full national independence or at least regional autonomy within given state structures made by a nationalist organization at the level of ethnic segment years (see Appendix A4).

6. We also use the variable separately in additional specifications in the Appendix.

Analyses and results

Our baseline specification is a difference-in-differences (DiD) regression estimated as two-way fixed effects (TWFE) linear probability model with the time-varying railway access dummy described above as treatment variable. The dependent variable is a combined indicator of national disintegration for all segment-years with either a successful secession, a territorial civil war onset, or a separatist claim for independence or regional autonomy. We multiply this outcome by 100 to increase readability and facilitate interpretation in terms of percentage points. All baseline models include unit fixed effects for ethnic segments and time fixed effects for either years or country-years – the latter control for the potential of regionally concentrated diffusion of secessionsm and other temporal shock and trends that equally affect all segments within a given country (e.g. Cunningham and Sawyer 2017). In addition, all models control for a count variable of past territorial civil wars since 1816 as well as peace year dummies for both civil war and nationalist claims to account for past secessionist mobilization and address concerns about reverse causation. The identifying assumption in this setup is parallel counterfactual trends, which we discuss in more detail below. Recent methodological contributions have highlighted problems with TWFE models when it comes to accommodating heterogeneous treatment effects across treatment cohorts and dynamically evolving effects after first treatment onset (e.g. Goodman-Bacon 2021; Callaway and Sant'Anna 2021). As our empirical setup resembles a multi-cohort difference-in-differences with staggered treatment adoption, we also implement a two-stage estimator recently proposed by Gardner (2021) and Liu, Wang, and Xu (2022). By imputing counterfactual outcomes for treated units based on a first-stage regression that only uses not-yet treated and/or never-treated units, the 2S-DiD approach alleviates most of the problems of conventional TWFE models.

Table 1 presents our main findings. Column 1 indicates that the probability of separatist claims, secessionist conflict fought in the name of a non-core ethnic segment, or successful se-

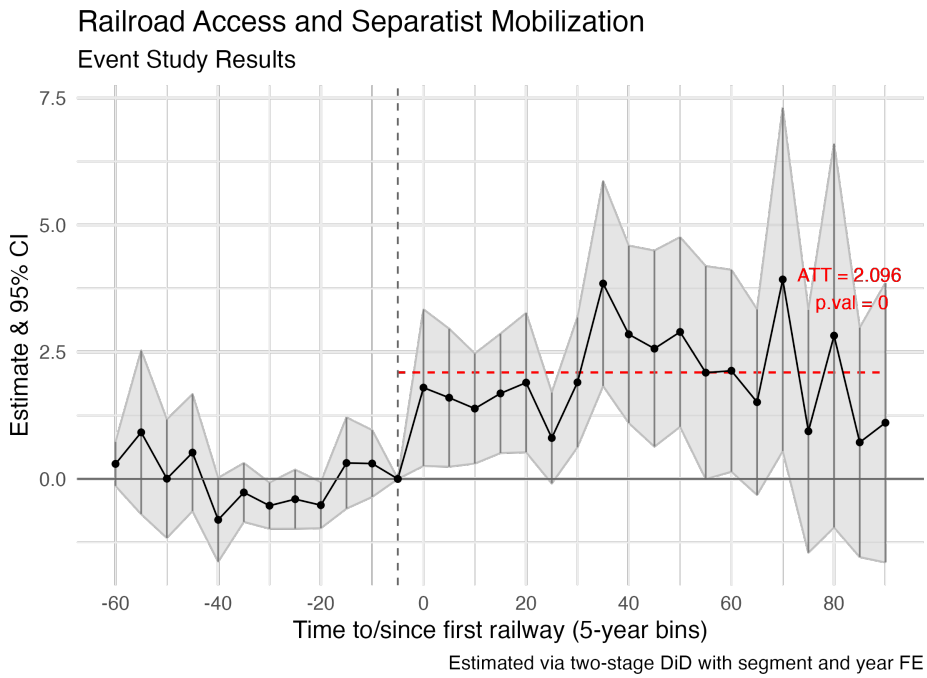
Table 1: Railroads and Separatism (1816-1945)

	100 × Secession, Terr. CW or Claim			
	(1)	(2)	(3)	(4)
Rails (Y/N)	1.486***	1.076**	2.096***	1.693***
	(0.352)	(0.341)	(0.493)	(0.447)
Segment FE	Yes	Yes	Yes	Yes
Year FE	Yes	No	Yes	No
Country-Year FE	No	Yes	No	Yes
Estimator	TWFE	TWFE	2S-DiD	2S-DiD
Mean DV	1.115	1.115	1.076	1.069
Observations	13 007	13 007	11 711	9818

Notes: The unit of analysis is the ethnic segment year. State-leading segments and segments smaller than 2000 sqkm dropped. Segment clustered standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

cession increases by 1.48 percentage points after the first railway arrives. This effect is substantively large and amounts to an almost 2.33-fold increase compared to the sample mean of 1.12 instances of separatist mobilization per 100 ethnic segment years. Column 2 replaces year with country-year fixed effects which reduces the estimated coefficient by 28%. Columns 3 and 4 replicate the analysis but rely on the two-stage DiD estimator developed by Gardner (2021) (see also Liu, Wang, and Xu 2022). Both specifications yield substantively larger estimates as their TWFE-based counterparts in Columns 1 and. Model 3 suggests a 195% increase from the sample mean, which drops to a 158% increase when replacing year with country-year fixed effects (Model 4). These results suggest that, on average and contrary to naive interpretations of modernization theory, railway access contributed to separatist mobilization rather than stronger national cohesion and political stability in ethnic minority areas.

Interpreting these findings as causal requires the assumption of parallel counterfactual trends. As counterfactual outcomes are by definition unobservable, we have to assume that, in the absence of treatment, treated units would have evolved similarly after treatment onset as not-yet-treated or never-treated control observations. While this assumption cannot be empirically verified, we can investigate trends before treatment onset to assess the plausibility of the identifying assumption.



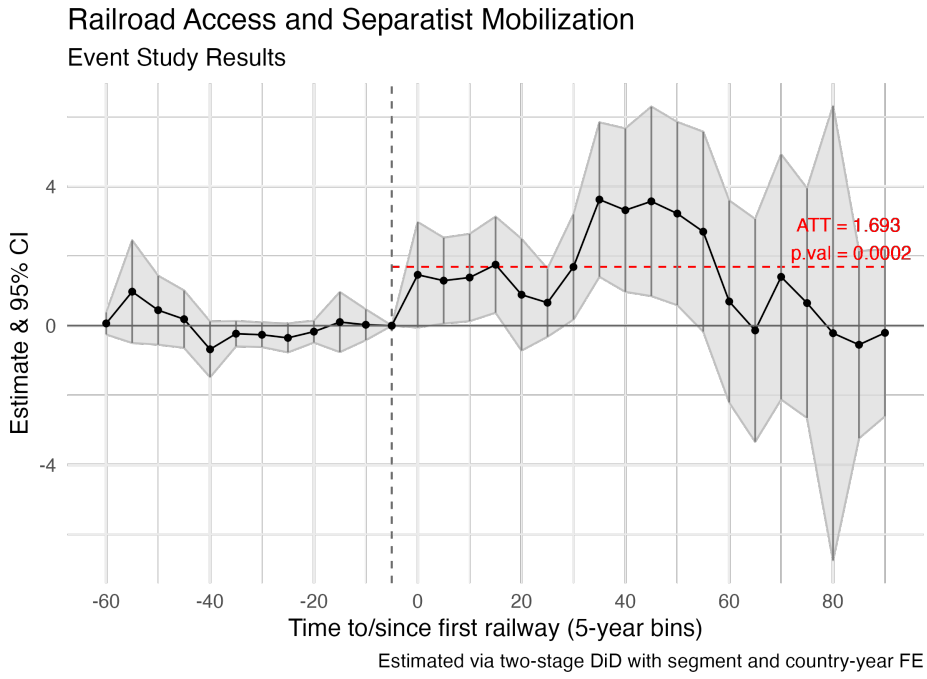


Figure 3: Event study plots
(ATT estimates based on Columns 3 and 4 in Table 1)

The top panel of Figure 3 plots coefficients and confidence intervals from a dynamic DiD specification ("event study") with segment and year fixed effects estimated via two-stage DiD. Instead of using a single post-treatment indicator, we now estimate coefficients for relative time-to-treatment bins. For each observation of all segments that eventually receive railways, we code (i) whether the segment year predates the first treatment year by more than 55, 51–55, 46–50, ..., 6-10, or 1–5 years and (ii) whether the first rail was built 0–4, 5–9, ..., 85–89, or more than 89 years ago. The first five-year bin before treatment onset is omitted and serves as the baseline category. The bottom panel shows results from an event study model using country-year instead of year fixed effects. Both plots reveal relatively parallel outcome trends between untreated and treated units in the periods before the latter receive their first railway line. The pre-trend dummy coefficients remain relatively close to zero and are jointly insignificant in both models. However, two pre-treatment dummies in the top panel are negative and significant at the 5% interval, which is not the case when using country-year fixed effects (bottom panel). The post-treatment dummies indicate an immediate increase in conflict risk after the first railway is built. The estimated treatment effects grow even larger approx. 35 years after the first railway and, if anything, diminish from the 50th post-treatment year onward (especially in the specification with country-year FE). These results clearly support Hypotheses 1 and cast some initial doubts on prominent integrationist mechanisms. Whether these mechanisms are irrelevant or still operative but, on average, outweighed by countervailing effects is a question we address below.

Robustness checks. As an alternative to controlling for past conflict in our baseline models, we run a robustness check that drops all ethnic segment-years that had a secessionist civil war or nationalist claims in the past. The results are summarized in Table A1 and show substantively smaller yet positive and significant treatment effects in all four models. In addition, we replicate our baseline results using a subsample that excludes all never-treated units. If ethnic

segments that never received a railway connection before 1945 are too small, rural, and peripheral to serve as valid comparison group for modernizing segments, their inclusion may reduce the credibility of parallel counterfactual trends and lead to biased conclusions. Appendix Table A2 summarizes our findings and shows, similar or, when using the two-stage DiD estimator, significantly larger treatment effects.

In the appendix to this paper, we disaggregate the outcome variable and report separate regressions for successful secessions, secessionist civil wars, and national independence or autonomy claims. The results in Tables A6–A7 suggest that our baseline findings are mainly driven by territorial civil wars and nationalist claims. That said, the estimated effects on the most extreme outcome of successful secession are positive and reach significance when estimated as two-stage DiD but substantively small and insignificant when using TWFE.

Table 2: Network Structure and Causal Mechanisms

	100 × Secession, Terr. CW or Claim			
	(1)	(2)	(3)	(4)
National Market Access	−0.143+			−0.001
	(0.083)			(0.075)
State Reach		−0.008**		−0.008**
		(0.003)		(0.003)
Internal Connectivity			0.015*	0.016*
			(0.007)	(0.007)
Segment FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Mean DV	1.131	1.115	1.115	1.131
Observations	12 643	13 007	13 007	12 643

Notes: The unit of analysis is the ethnic segment year. State-leading segments and segments smaller than 2000 sqkm dropped. Segment clustered standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Exploring causal mechanisms. In the next part of the analysis, we attempt to separate the three mechanisms through which railway construction may affect center-periphery bargaining and separatist mobilization as outlined in the theory. To that end, we compute railway-based proxies for (1) segments' economic market access as their average travel time towards large cities,

(2) local state's reach as their average travel time to the capital, and (3) their internal connectivity as the average travel time among their inhabitant.[7] Table 2 shows TWFE models of separatism where these variables replace our baseline railway dummy variable.

All coefficient estimates point in the expected direction and, with the exception of *National Market Access* in Columns 1 and 4, reach conventional significance levels. In line with top-down mechanisms of state-sponsored nation building, better links to the national capital reduce the likelihood of separatist mobilization as predicted by Hypothesis H2b. The effect of internal connectivity (M3 in Figure 1) points towards a higher capacity of local elites and populations to organize collective action against the state, which is consistent with Hypothesis H2c. The negative and borderline significant coefficient of *National Market Access* turns small, positive, and insignificant when also including state reach, which suggests that most of the negative effect in the first model seems to be driven by better connections to the capital city. Taken together, these results provide stronger support for the more political and mobilization-related mechanisms M2 and M3 than for bottom-up nation building via market integration and social communication (M1). Another possibility is that increasingly integrated national railroad networks exert heterogeneous effects across different contexts and that, on average, integrative and disintegrative responses balance each other out. The fact that our baseline analysis shows positive effects of the first railway link in a segment on separatist mobilization may thus be due to peripheral connections in historical Europe mainly strengthening local ties rather than effectively boosting state capacity or integrating national markets.

That said, these findings by no means imply that reactive nationalism and local resistance against direct rule are irrelevant. One interpretation is that such resistance needs to occur before it's too late or, more specifically, after railway access and internal connectivity improve local mobilization capacity but before the state thoroughly penetrates or even assimilates peripheral populations (Deutsch 1953). It is also possible that a more selective indicator for culturally

7. See Appendix A3.

distinctive direct rule of "nationalizing" states (Brubaker 1996) that goes beyond mere travel time would yield different results.

Testing our conditional hypotheses. To test the conditional hypotheses described above, we replicate the baseline model from Column 1 in Table 1 but now interact the railway access dummy with segment- or country-year moderating variables. Marginal effect plots of all interaction models are presented in the Appendix (Figures A4-A9) along with the binning estimates recently proposed by Hainmueller, Mummolo, and Xu (2019). Column 1 in Table 3 tests whether the destabilizing effects of rails is stronger in ethnic segments that are culturally more distinct from the state-leading group (H3a). We calculate linguistic distance from the core group by matching the ethnic categories from our maps to the Ethnologue language tree. Interacting our rail treatment with linguistic distance yields a positive but statistically insignificant coefficient (Model 1 in Table 3). One interpretation of this non-result is that conditional on some cultural difference, group-level politicization and mobilization processes are more important than cultural distance.

Model 2 interacts the rail indicator with the country-year-level population share of the dominant national core group. Consistent with Hypothesis 3b, the interaction coefficient is negative and significant suggesting local railways are particularly likely to spur nationalist independence campaigns in countries with relatively small ruling groups. However, the binning plot in Figure A5 suggests that the significant linear interaction term is likely due to small number of cases with particularly small core groups. The binning coefficients show that there are no significantly different effects at typically low, intermediate, and high values of the respective national core group's population share. Model 3 tests our argument about the opportunities of non-core groups to mobilize for secession. The treatment and interaction coefficient reveal that railways mainly spur territorial conflict in demographically large ethnic segments, in line with hypotheses H3c.

Table 3: Separatism: Interaction Models

	100 × Secession, Terr. CW or Claim					
	(1)	(2)	(3)	(4)	(5)	(6)
Rails (Y/N)	0.493	3.484***	−10.291***	16.278***	2.761***	2.331***
	(0.611)	(0.817)	(2.103)	(4.308)	(0.629)	(0.506)
Rails × Ling. Dist to Core	1.373+					
	(0.820)					
Pop. Share Core Group		0.505				
		(2.328)				
Rails × Pop. Share Core		−3.834**				
		(1.202)				
Group Population (log)			−0.141			
			(0.255)			
Rails × Group Pop.			0.908***			
			(0.170)			
GDP per capita (log)				0.922		
				(1.033)		
Rails × GDP p.c.				−1.882***		
				(0.533)		
Fiscal Capacity (VDEM)					−0.229	
					(0.254)	
Rails × Fiscal Cap.					−0.912**	
					(0.339)	
Liberal Democracy (VDEM)						0.141
						(1.309)
Rails × Lib. Dem.						−2.317**
						(0.818)
Segment FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean DV	1.115	1.115	1.115	1.134	1.134	1.146
Observations	13 007	13 007	13 007	12 788	12 788	12 649

Notes: The unit of analysis is the ethnic segment year. State-leading segments and segments smaller than 2000 sqkm dropped. Segment clustered standard errors in parentheses. $+ p < 0.1$, $* p < 0.05$, $** p < 0.01$, $*** p < 0.001$.

To test H3d and H3e, we rely on historical VDEM data on per capita GDP and fiscal capacity. The negative and significant interaction term with per capita income suggests that our findings are driven by relatively poor and arguably less industrialized country-years in the sample, thus confirming H2d. The linear interaction term with fiscal capacity is negative and significant and the binning estimates in Figure A8 suggest that the effect of railway access is significantly larger at typically low values of fiscal capacity than at typical medium or high values, consistent with H3e. Last but not least, the interaction term with the VDEM liberal democracy score is negative

and significant (Column 6). However, the binning estimates in Appendix Figure A9 reveal that, if anything, the effect is highest at intermediate values of liberal democracy. While the rail effect at typically high democracy values is significantly smaller than at intermediate levels, it is not significantly lower than at the lowest levels of democracy.

Accounting for reverse causation. Our DiD research design allows us to rule out most sources of unobserved variance. Still, the very decision to build railway lines in certain segments over others at a given time could be due to reverse causality. Our brief overview of motivations behind railroad construction above shows that, particularly in the Russian case, such projects could have been at least partly influenced by internal security considerations.

In order to address this potential caveat, we implement an instrumental variables strategy based on simulated, time-varying railroad networks. Our simulation procedure starts from the intuition that, in the absence of local policing or external military goals, state-provided railroad networks would aim to maximize the connectedness within a country's population. We furthermore assume that each state has a fixed budget of railroad kilometers to build every year. On that basis, we build a planar graph that covers all of Europe. This network connects each cell of a population raster with a resolution of .5 decimal degrees (≈50km at the equator and less as one moves North) to its eight nearest neighbors.

Up to 1833, the network only consists of foot- and carriage paths on which one can travel 6km/h. Our simulation algorithm, described in full detail in Appendix A5, now "builds" the observed railroad mileage for every consecutive year as upgrades to these baseline paths, increasing the allowed travel speed to 60km/h for every edge transformed into a railroad line. In doing so, the algorithm heuristically places railroad lines such that they minimize the average travel time between any two inhabitants of the same country. The crucial input to this algorithm is a time-*in*variant estimated population grid for the year 1830 from Goldewijk, Beusen, and Janssen 2010.

The simulation is driven by the spatial interaction of four factors. First, the continental population distribution in 1830 ensures that most rails are build around and between population centers. Second, (changing) country borders affect which areas are central or peripheral to states' networks. Third, states observed annual railroad budget affects the evolution of the railroad over time. And fourth, the stock of simulated railroads build in previous years affects where the next set of lines are being built.

The exclusion restriction for the use of the simulated railroad networks as instruments for the observed ones assumes that the simulation results only affect separatism through the building of railroad infrastructure and are not systematically affected by proximate causes of conflict. While the first requirement is relatively unproblematic, there may be a risk of omitted variable bias, in particular where large populations cause railway construction and conflict. In response, our segment fixed effects control for any time-invariant factors at the level of segments such as their population, geography, and location within a state. Furthermore, year fixed effects account for the overall trend in the expansion of railroad networks. In the Appendix, we additionally show that results are robust to adding country-year fixed effects which account for the timing of state-specific railroad investments and border changes which affect the railroad simulation and potentially conflict. As a result, the instrumented railway measure only pick up effects on conflict through the spatio-temporal sequencing of railroad construction as modeled by our simulation, namely that some ethnic settlement areas are likely to be connected to the network earlier than others.

We aggregate the simulated railroads onto our ethnic segments and use the resulting, time-varying segment-level measure of railroad exposure as an instrumental variable for observed railroad exposure in a two-way fixed effect setup. Table 4 summarizes the relevant models. Column 1 presents the first stage of the IV estimation and shows that the simulation-based instrument is strongly predictive of actual railway construction within peripheral ethnic segments. The relevant F statistic of 39 is well above conventional thresholds of instrument strength. Column 2

Table 4: Instrumenting Railroad Access

	Rails (Y/N)	100 × Separatism		
	First Stage	OLS	Reduced Form	Second Stage
Rails (Y/N, simulated)	0.335***		0.785*	
	(0.054)		(0.321)	
Rails (Y/N)		1.514***		
		(0.375)		
Rails (Y/N, instrumented)				2.341*
				(0.975)
Segment FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
First Stage F	38.746			38.746
Mean DV	0.512	1.115	1.115	1.115
Observations	13 007	13 007	13 007	13 007

Notes: The unit of analysis is the ethnic segment year. State-leading segments and segments smaller than 2000 sqkm dropped. Segment clustered standard errors in parentheses. $+ p < 0.1$, $* p < 0.05$, $** p < 0.01$, $*** p < 0.001$.

replicates our TWFE baseline to allow direct comparisons with Columns 3 and 4. Model 3 is a reduced form regression of separatism on the simulated instrument, whereas Column 4 shows the second-stage estimate of instrumented rail access. Both relevant coefficients are positive and significant at the 5% level. The IV coefficient is 55% larger than the baseline TWFE effect in Column 2 but standard errors increase as it is to be expected in IV models. Replacing year with country-year fixed effects, if anything, leads to stronger results as shown in Appendix Table A4. These findings increase our confidence that the estimated effects are not merely reflecting reverse causation resulting from strategic railway construction in areas threatened by separatism.

Conclusion

Modern transportation infrastructure is conventionally seen as having strengthened 19th and 20th century European state- and nation-building. Expanding railway networks boosted centralizing states' infrastructural power and enabled increasingly direct forms of governance. At the same time, they spurred economic change, urbanization, and social contact over increasing distance.

Extrapolating from Weber's (1976) case study of nationalizing France, many contemporary social scientists expect these changes to have strengthened national attachments and cohesion well beyond the French case. In a classical study, Connor (1972) criticized "the very common habit of describing Western Europe as though it were composed of fully integrated states" arguing that "Western Europe is therefore held up as a model of something it is not, as proof that something can be achieved elsewhere that is in fact far from achieved there."

In this paper, we have shown that, if anything, railway construction in ethnic minority regions tended to threaten the integrity of existing European states and empires. Our difference-in-differences analyses suggest that secessionist conflict, independence claims, and successful secession became more likely after territories inhabited by non-leading ethnic groups were connected to the state's railroad network. Using network-derived continuous proxies of railroad expansion, we show that separatism becomes more likely where railroads facilitate mobilization by improving internal connectivity of peripheral ethnic regions and less likely where it brings such regions closer to the capital. National market access, approximated as connectivity to major cities within the country does not seem to make a difference.

In light of these findings, the French experience appears much more as an outlier than a paradigmatic case for the link between modernization and nationalism in historical Europe. Our conditional hypotheses and tests reveal some structural dimensions which confirm that modernization processes did not yield national integration in most multi-ethnic states, especially in Eastern Europe. Relatively small population shares of state-leading groups, comparatively weak levels of historically inherited state capacity, low per capita incomes as well as large minority groups posed formidable challenges for state centralization and top-down nation-building.

Finally, it should be recalled that railway construction was only one, though arguably the most important, vector of modernization in Europe from the 19th century through the mid-20th century, especially since many other mechanisms depended on rail infrastructure. In this sense, the current study contributes to a broader literature that analyzes national integration

or disintegration through various means of social communication and mechanisms of identity formation, such as telegraph lines, road networks, mass education and mass media. In addition, there is a quickly developing research agenda showing how mobilization processes around the world are influenced by more recent technologies, such as broadcasting (Warren 2014), cell-phone technology (Shapiro and Weidmann 2015) or social media (Weidmann 2015; Gohdes 2020). While our study serves as a reminder that technological advances sometimes have disintegrating effects, careful empirical research will have to be conducted before applying our findings to settings beyond the classical cases of European nation building.

References

Abdelgadir, Aala, and Vasiliki Fouka. 2020. "Political secularism and Muslim integration in the West: Assessing the effects of the French headscarf ban." *American Political Science Review* 114 (3): 707–723.

Aghion, Philippe, Xavier Jaravel, Torsten Persson, and Dorothée Rouzet. 2019. "Education and military rivalry." *Journal of the European Economic Association* 17 (2): 376–412.

Aidt, Toke, Gabriel Leon-Ablan, and Max Satchell. 2022. "The social dynamics of collective action: evidence from the diffusion of the Swing riots, 1830–1831." *The Journal of Politics* 84 (1): 209–225.

Alesina, Alberto, Paola Giuliano, and Bryony Reich. 2021. "Nation-building and education." *The Economic Journal* 131 (638): 2273–2303.

Alesina, Alberto, and Bryony Reich. 2015. "Nation building." *Working Paper*.

Alvarez-Palau, Eduard J., Alfonso Díez-Minguela, and Jordi Martí-Henneberg. 2021. "Railroad Integration and Uneven Development on the European Periphery, 1870-1910." *Social Science HIstory* 45:261–289.

Anderson, Benedict O G. 1983. *Imagined Communities: Reflections on the Origin and Spread of Nationalism.* Verso.

Ansell, Ben W., and Johannes Lindvall. 2021. *Inward Conquest: The Political Origins of Modern Public Services.* Cambridge: Cambridge University Press. Book.

Assouad, Lydia. 2020. "Charismatic leaders and nation building." *Unpublished Working Paper.*

Balcells, Laia. 2013. "Mass schooling and Catalan nationalism." *Nationalism and Ethnic Politics* 19 (4): 467–486.

Balcells, Laia, Lesley-Ann Daniels, and Alexander Kuo. n.d. "The 'weight' of territorial issues: Evidence from Catalonia, Scotland, and Northern Ireland." *Comparative Politics, forthcoming.*

Bates, Robert H. 1983. "Modernization, Ethnic Competition, and the Rationality of Politics in Contemporary Africa." In *State Versus Ethnic Claims: African Policy Dilemmas,* edited by Donald Rothchild and Victor A Olorunsola, 152–171. Boulder, CO: Westview Press.

Berger, Thor. 2019. "Railroads and rural industrialization: Evidence from a historical policy experiment." *Explorations in Economic History* 74:101–277.

Berger, Thor, and Kerstin Enflo. 2017. "Locomotives of local growth: The short-and long-term impact of railroads in Sweden." *Journal of Urban Economics* 98:124–138.

Blanc, Guillaume, and Masahiro Kubo. 2022. "Schools, Language, and Nations: Evidence From a Natural Experiment in France." *Unpublished Working Paper.*

Bogart, Dan. 2009. "Nationalizations and the Development of Transport Systems: Cross-Country Evidence from Railroad Networks, 1860–1912." *The Journal of Economic History* 69 (1): 202–237.

Breuilly, John. 1982. *Nationalism and the State.* Manchester, UK: Manchester University Press.

Brooke, Steven, and Neil Ketchley. 2018. "Social and institutional origins of political Islam." *American Political Science Review* 112 (2): 376–394.

Brubaker, Rogers. 1996. *Nationalism Reframed: Nationhood and the National Question in the New Europe*. Cambridge: Cambridge University Press. Book.

Burbank, Jane, and Frederick Cooper. 2010. *Empires in World History: Power and the Politics of Difference*. Princeton University Press.

Buzan, Barry, and George Lawson. 2017. *The Global Transformation: History, Modernity and the Making of International Relations*. Cambridge: Cambridge University Press. Book.

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with multiple time periods." Themed Issue: Treatment Effect 1, *Journal of Econometrics* 225 (2): 200–230.

Cermeño, Alexandra L, Kerstin Enflo, and Johannes Lindvall. 2022. "Railroads and reform: How trains strengthened the nation state." *British Journal of Political Science* 52 (2): 715–735.

Choi, Donghyun Danny, Benjamin Laughlin, and Anna E Schultz. 2021. "Mobile communication technology and national identity in Sub-Saharan Africa."

Christensen, Darin, and Francisco Garfias. 2018. "Can you hear me now? How communication technology affects protest and repression." *Quarterly Journal of Political Science* 13 (1): 89.

Cinnirella, Francesco, and Ruth Schueler. 2018. "Nation building: The role of central spending in education." *Explorations in Economic History* 67:18–39.

Connelly, John. 2020. *From Peoples into Nations: A History of Eastern Europe*. Princeton, NJ: Princeton University Press. Book.

Connor, Walker. 1972. "Nation-Building or Nation-Destroying." *World Politics* 24:319–355.

Cunningham, Kathleen Gallagher, and Katherine Sawyer. 2017. "Is self-determination contagious? A spatial analysis of the spread of self-determination claims." *International Organization* 71 (3): 585–604.

Darden, Keith, and Harris Mylonas. 2016. "Threats to territorial integrity, national mass schooling, and linguistic commonality." *Comparative Political Studies* 49 (11): 1446–1479.

Deutsch, Karl W. 1953. *Nationalism and Social Communication: An Inquiry into the Foundations of Nationality*. MIT Press.

Donaldson, Dave. 2018. "Railroads of the Raj: Estimating the impact of transportation infrastructure." *American Economic Review* 108 (4-5): 899–934.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American economic growth: A "market access" approach." *The Quarterly Journal of Economics* 131 (2): 799–858.

Eifert, Benn, Edward Miguel, and Daniel N Posner. 2010. "Political Competition and Ethnic Identification in Africa." *American Journal of Political Science* 54 (2): 494–510.

Elliott, J. H. 1992. "A Europe of Composite Monarchies." *Past and Present,* no. 137, 48–71.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2020. "Social media and protest participation: Evidence from Russia." *Econometrica* 88 (4): 1479–1514.

Fishlow, Albert. 1965. *American Railroads and the Transformation of the Ante-bellum Economy*. Vol. 127. Cambridge, MA: Harvard University Press.

Fouka, Vasiliki. 2020. "Backlash: The unintended effects of language prohibition in US schools after World War I." *The Review of Economic Studies* 87 (1): 204–239.

García-Jimeno, Camilo, Angel Iglesias, and Pinar Yildirim. 2022. "Information Networks and Collective Action: Evidence from the Women's Temperance Crusade." *American Economic Review* 112 (1): 41–80.

Gardner, John. 2021. "Two-stage differences in differences." *Working paper.*

Garfias, Francisco, and Emily A Sellars. 2021a. "From conquest to centralization: Domestic conflict and the transition to direct rule." *The Journal of Politics* 83 (3): 992–1009.

———. 2021b. "When State Building Backfires: Elite Coordination and Popular Grievance in Rebellion." *American Journal of Political Science.*

Gellner, Ernest. 1964. *Thought and Change.* London: Widenfeld & Nicolson. Book.

———. 1983. *Nations and Nationalism.* Cornell University Press.

Gleditsch, Kristian Skrede. 2004. "A revised list of wars between and within independent states, 1816-2002." *International Interactions* 30 (3): 231–262.

Gleditsch, Kristian Skrede, and Michael D Ward. 1999. "Interstate system membership: A revised list of the independent states since 1816." *International Interactions* 25 (4): 393–413.

Gohdes, Anita R. 2020. "Repression Technology: Internet Accessibility and State Violence." *American Journal of Political Science* 64 (3): 488–503.

Goldewijk, Kees Klein, Arthur Beusen, and Peter Janssen. 2010. "Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1." *The Holocene* 2010 (1): 1–9.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." Themed Issue: Treatment Effect 1, *Journal of Econometrics* 225 (2): 254–277.

Granovetter, Mark. 1978. "Threshold models of collective behavior." *American Journal of Sociology* 83 (6): 1420–1443.

Green, Elliott D. 2022. *Industrialization and Assimilation: Understanding Ethnic Change in the Modern World.* New York: Cambridge University Press.

Gutkas, Kurt, and Ernst Bruckmüller, eds. 1989. *Verkehrswege und Eisenbahnen: Beiträge zur Verkehrsgeschichte Österreichs aus Anlass des Jubiläums «150 Jahre Dampfeisenbahn in Österreich».* Schriften des Instituts für Österreichkunde. Wien: Wien: Österreichischer Bundesverlag.

Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu. 2019. "How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice." *Political Analysis* 27 (2): 163–192.

Hechter, Michael. 1977. *Internal colonialism: The Celtic fringe in British national development, 1536-1966.* Vol. 197. Univ of California Press.

———. 2000. *Containing Nationalism.* New York: Oxford University Press.

———. 2013. *Alien Rule.* New York: Cambridge University Press.

Hierro, Maria Jose, and Didac Queralt. 2021. "The divide over independence: Explaining preferences for secession in an advanced open economy." *American Journal of Political Science* 65 (2): 422–442.

Hintze, Otto. 1975. "Military Organization and the Organization of the State," edited by Felix Gilbert, vol. The Historical Essays of Otto Hintze. New York: Oxford University Press.

Hobsbawm, Eric. J. 1990. *Nations and Nationalism Since 1780.* Cambridge: Cambridge University Press. Book.

Hornung, Erik. 2015. "Railroads and growth in Prussia." *Journal of the European Economic Association* 13 (4): 699–736.

Huntington, Samuel P. 1968. *Political Order in Changing Societies.* New Haven: Yale University Press.

Keller, Wolfgang, and Carol H Shiue. 2008. "Institutions, Technology, and Trade." *NBER Working Paper,* no. w13913.

Köster, Burkhard. 1999. *Militär und Eisenbahn in der Habsburgermonarchie 1825-1859.* München: R. Olderburg Verlag. Book.

Kuran, Timur. 1992. "Now out of never: the element of surprise in the East European revolution of 1989." *World Politics* 44:7–48.

Lewis, M. Paul, ed. 2009. *Ethnologue: Languages of the world.* Vol. 16. SIL International Dallas, TX.

Liu, Licheng, Ye Wang, and Yiqing Xu. 2022. "A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data." *American Journal of Political Science* Early View.

Maier, Charles S. 2016. *Once Within Borders: Territories of Power, Wealth, and Belonging since 1500.* Harvard University Press.

Manacorda, Marco, and Andrea Tesei. 2020. "Liberation technology: Mobile phones and political mobilization in Africa." *Econometrica* 88 (2): 533–567.

Mann, Michael. 1993. *The Sources of Social Power. Volume 2: The Rise of Classes and Nation-States, 1760–1914.* New York: Cambridge University Press.

Melander, Eric. 2021. "Transportation Technology, Individual Mobility and Social Mobilisation."

Miller, Alexei. 2015. "The Romanov Empire and the Russian Nation." In *Nationalizing Empires,* edited by Stefan Berger and Alexei Miller. Budapest: Central European University Press.

Motyl, Alexander J. 1997. "Thinking about Empire." In *After Empire: Multiethnic Societies and Nation Building - The Soviet Union And The Russian, Ottoman, And Habsburg Empires,* edited by Karen Barkey, 19–29. New York: Routledge.

Müller-Crepon, Carl, Philipp Hunziker, and Lars-Erik Cederman. 2021. "Roads to rule, roads to rebel: relational state capacity and conflict in Africa." *Journal of Conflict Resolution* 65 (2-3): 563–590.

Mylonas, Harris. 2012. *The politics of nation-building: Making co-nationals, refugees, and minorities.* Cambridge University Press.

Osterhammel, Jürgen. 2014. *The Transformation of the World: A Global History of the Nineteenth Century.* Princeton, NJ: Princeton University Press. Book.

Paglayan, Agustina S. 2021. "The non-democratic roots of mass education: evidence from 200 years." *American Political Science Review* 115 (1): 179–198.

———. 2022. "Education or indoctrination? The violent origins of public school systems in an era of state-building." *American Political Science Review,* Online First.

Pierskalla, Jan H, and Florian M Hollenbach. 2013. "Technology and collective action: The effect of cell phone coverage on political violence in Africa." *American Political Science Review* 107 (2): 207–224.

Posen, Barry R. 1993. "Nationalism, the Mass Army, and Military Power." *International Security* 18 (2): 80–124.

Reed, Frank. 2008. *Centennia Historical Atlas.*

Rieber, Alfred. 2014. *The Struggle for the Eurasian Borderlands: From the Rise of Early Modern Empires to the End of the First World War.* Cambridge: Cambridge University Press. Book.

Robinson, Amanda Lea. 2014. "National versus ethnic identification in Africa: Modernization, colonial legacy, and the origins of territorial nationalism." *World Politics* 66 (4): 709–746.

Rohner, Dominic, and Ekaterina Zhuravskaya. 2023. *Nation Building: Big Lessons from Successes and Failures.* London: CEPR Press. Book.

Rostow, Walt Whitman. 1960. *The stages of economic growth: A non-communist manifesto.* New York: Cambridge university press.

Sarkees, Meredieth Reid, and Frank Wayman. 2010. *Resort to War: 1816–2007. Correlates of War.* Washington DC: CQ Press.

Schenk, Frithjof Benjamin Schenk. 2011. "Master Imperial Space? The Ambivalent Impact of Railway-Building in Tsarist Russia." In *Comparing Empires: Encounters and Transfers in the Long Nineteenth Century,* edited by Jörn Lonard and Ulrike von Hirschhausen. Göttingen: Vandenhoeck & Ruprecht.

Schram, Albert. 1997. *Railways and the Formation of the Italian State in the Nineteenth Century.* Cambridge: Cambridge University Press. Book.

Schvitz, Guy, Luc Girardin, Seraina Ruegger, Nils B. Weidmann, Lars-Erik Cederman, and Kristian S. Gleditsch. 2022. "Mapping the International System, 1886-2019: The CShapes 2.0 Dataset." *Journal of Conflict Resolution* 66 (1): 144–161.

Segal, Zef. 2016. "Regionalism and Nationalism in the Railway Cartography of Mid-Nineteenth Century Germany." *Imago Mundi* 68 (1): 46–61.

Shapiro, Jacob N., and Nils B. Weidmann. 2015. "Is the Phone Mightier Than the Sword? Cellphones and Insurgent Violence in Iraq." *International Organization* 69 (2): 247–274.

Shesterinina, Anastasia. 2016. "Collective threat framing and mobilization in civil war." *American Political Science Review* 110 (3): 411–427.

Siroky, David S, Sean Mueller, and Michael Hechter. 2016. "Center-periphery bargaining in the age of democracy." *Swiss Political Science Review* 22 (4): 439–453.

Tilly, Charles. 1994. "States and nationalism in Europe 1492-1992." *Theory and society,* 131–146.

Tir, Jaroslav, Philip Schafer, Paul F Diehl, and Gary Goertz. 1998. "Territorial changes, 1816–1996: Procedures and data." *Conflict Management and Peace Science* 16 (1): 89–97.

Trew, Alex. 2020. “Endogenous Infrastructure Development and Spatial Takeoff in the First Industrial Revolution.” *American Economic Journal: Macroeconomics* 12, no. 2 (April): 44–93.

Warren, T. Camber. 2014. “Not by the Sword Alone: Soft Power, Mass Media, and the Production of State Sovereignty.” *International Organization* 68 (1): 111–141.

Weber, Eugen. 1976. *Peasants into Frenchmen: The Modernization of Rural France, 1870-1914.* Palo Alto, CA: Stanford University Press.

Weidmann, Nils B. 2015. “Communication, technology, and political conflict: Introduction to the special issue.” *Journal of Peace Research* 52 (3): 263–268.

Wimmer, Andreas. 2018. *Nation Building: Why Some Countries Come Together While Others Fall Apart.* Princeton, NJ: Princeton University Press. Book.

Wimmer, Andreas, and Yuval Feinstein. 2010. “The Rise of the Nation-State across the World, 1816 to 2001.” *American Sociological Review* 75 (5): 764–790.

Wucherpfennig, Julian, Nils W. Metternich, Lars-Erik Cederman, and Kristian Skrede Gleditsch. 2012. “Ethnicity, the State, and the Duration of Civil War.” *World Politics* 64 (1): 79–115.

Supplementary Materials (Online)

A1 Ethnic settlement data

The HEG dataset covering ethnic settlement area is based on a candidate set of approximately 200 historical ethnic maps compiled from online map collections and leading libraries such as the British Library, Library of Congress, and the Blbliothèque Nationale de France. From this candidate set, we selected 73 high-quality maps with (a) high geographic resolution, (b) broad spatial coverage (i.e. depicting large subregions or the entirety of Europe), (c) authors of varying nationality, and (d) no obvious political biases.[8]

Practically all ethnic categories appearing on our maps refer to linguistic rather than religious or regional ethnic identity markers. That said, some maps differ in the level of linguistic granularity they encode and therefore need to be standardized for our purposes. To address this "grouping problem" of European ethnolinguistic identities, we match all raw linguistic map labels to the Ethnologue language tree (Lewis 2009) and construct a time-invariant master list of relevant ethnolinguistic groups by subsuming linguistically closely related labels from different

8. The publication of these maps range from the 1850s to 2019. For the present project, we restrict ourselves to the period 1816-1945.

maps under the linguistic node that occurs on the majority of maps that depict the respective language family.[9]

To get at temporal variation in specific groups' settlement areas, we combine the publication date of individual maps as well as hand-coded secondary data on the relatively few periods of large-scale ethnic change due to forced resettlement, genocide, or mass migrations. This information is used to code, for each group on our ethnic master list, the maps that are valid for a specific sub-period between 1816 and 1945.[10]

Finally, we draw on all maps belonging to a specific group-time period combination to construct a best-guess settlement polygon. Figure A1 illustrates this procedure for the Hungarian map period before WWII. The first step is to overlay the digitized multipolygons of all 12 maps that show the Hungarians. Second, we rasterize these polygons and calculate, for each raster cell, the share of maps that encode it as populated by Hungarians. The third and final step applies a 0.5 cutoff rule to construct a best-guess polygon that contains all cells that at least six maps regard as populated by Hungarians. These best-guess polygons may, of course, overlap, which indicates mixed settlements.

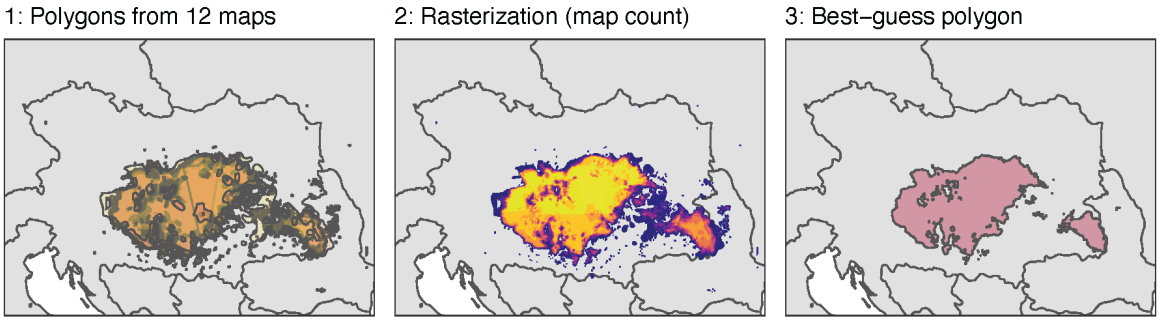


Figure A1: Constructing ethnic buest-guess polygons: Hungarian example

A2 Validation of railway data

We validate the quality of the main spatial railway data using a set of hand-geocoded historical railway maps for Austria Hungary. We collected a total of 12 maps for 1855, 1864, 1869, 1870,

9. If, for example, two maps contain the Bavarian dialect while twenty maps depict Germans, the Germans are listed as relevant group and subsume all dialects. In other cases, more disaggregate categories are chosen. Croats, Serbians, and Bosniaks appear on many more maps than does the aggregate South Slavic language family.

10. To address concerns that accurately reflecting temporal change in ethnic settlements comes at the cost of introducing endogeneity problems to our analyses, we run robustness checks only relying on the earliest available maps (NOTE: output tables and add to Appendix).

1876, 1881, 1884, 1885, 1991, 1995, 1901 from the Rumsay historical map collection. Each map is georeferenced and its railway lines drawn with the help of contemporary OpenStreetMaps railroad data. This helps improving the precision of lines, and only in few cases additional lines needed to be drawn by hand.

As a first visual validation exercise we compare the main railway dataset (henceforth RShapes) to the hand-drawn Austro-Hungarian lines. The left-hand plots in Figure A2 overlay the two sets of lines for 4 years: 1855, 1870, 1884, and 1901. The comparisons suggest that RShapes correctly identifies the main rail lines in Austria-Hungary. If anything, it somewhat underestimates the density of rail connections, especially in 1901.

As a second step we sample points on the hand-drawn lines circa every 5 kilometers and estimate the average distance of these points to the nearest RShapes line, as well as computing the share of points that lie within a 10-kilometer buffer around RShapes lines. These two metrics should give a quantitative measure of the two line sets' agreement. The right-hand plots in Figure A2 describe the points and buffers. It is clear that the largest share of points are nested within the RShapes buffer areas.

Plots in Figure A3 plot the average distance between rail lines and the share of points on the hand-drawn lines that fall within 10 kilometers from RShapes lines. Both statistics show fairly low error rates. In particular, the share of points within the 10-kilometer buffer shows fairly high consistency over all observed years, as more than 75 percent of all points are within the buffer area.

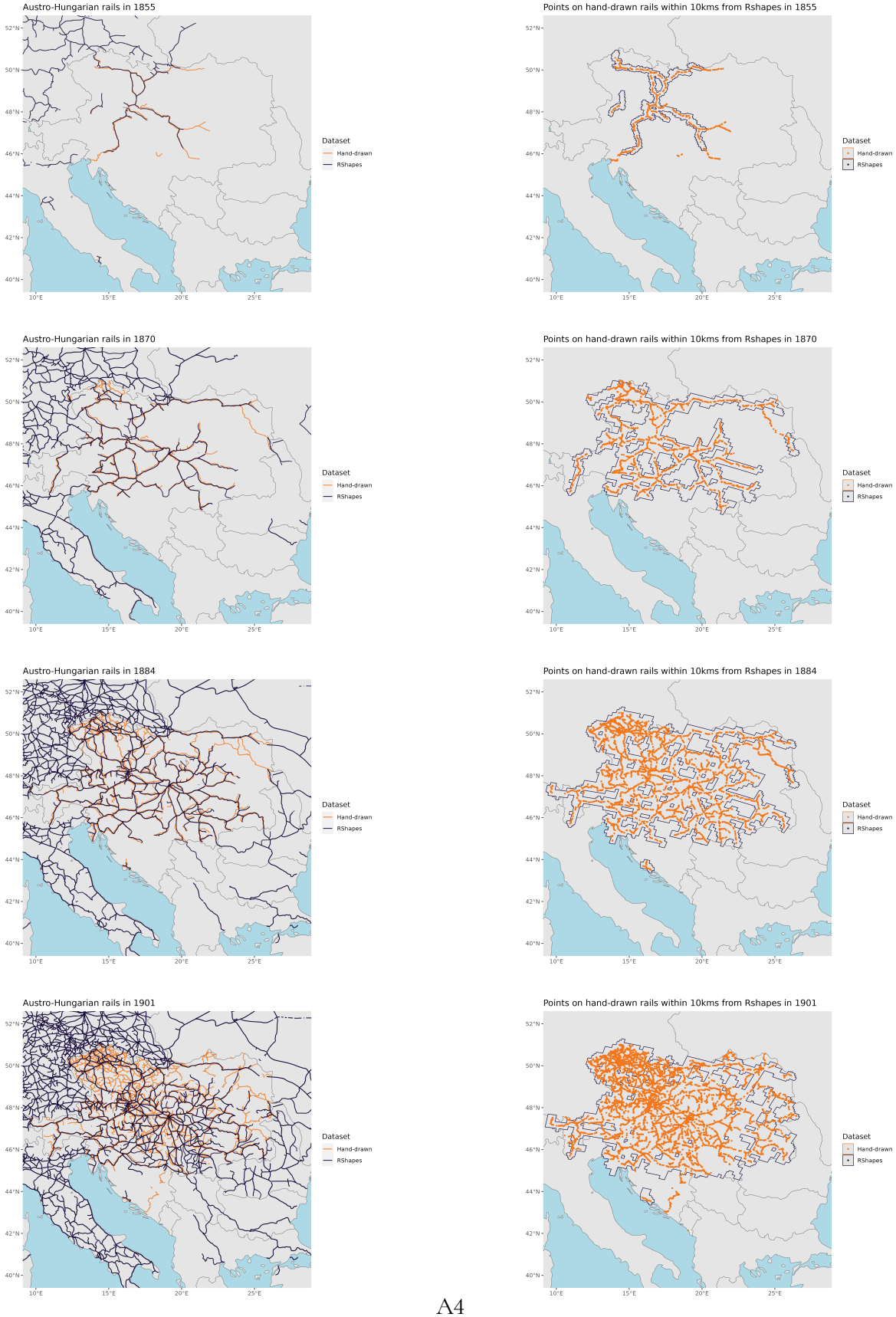


Figure A2: Comparison of RShapes and Austro-Hungarian railway data.

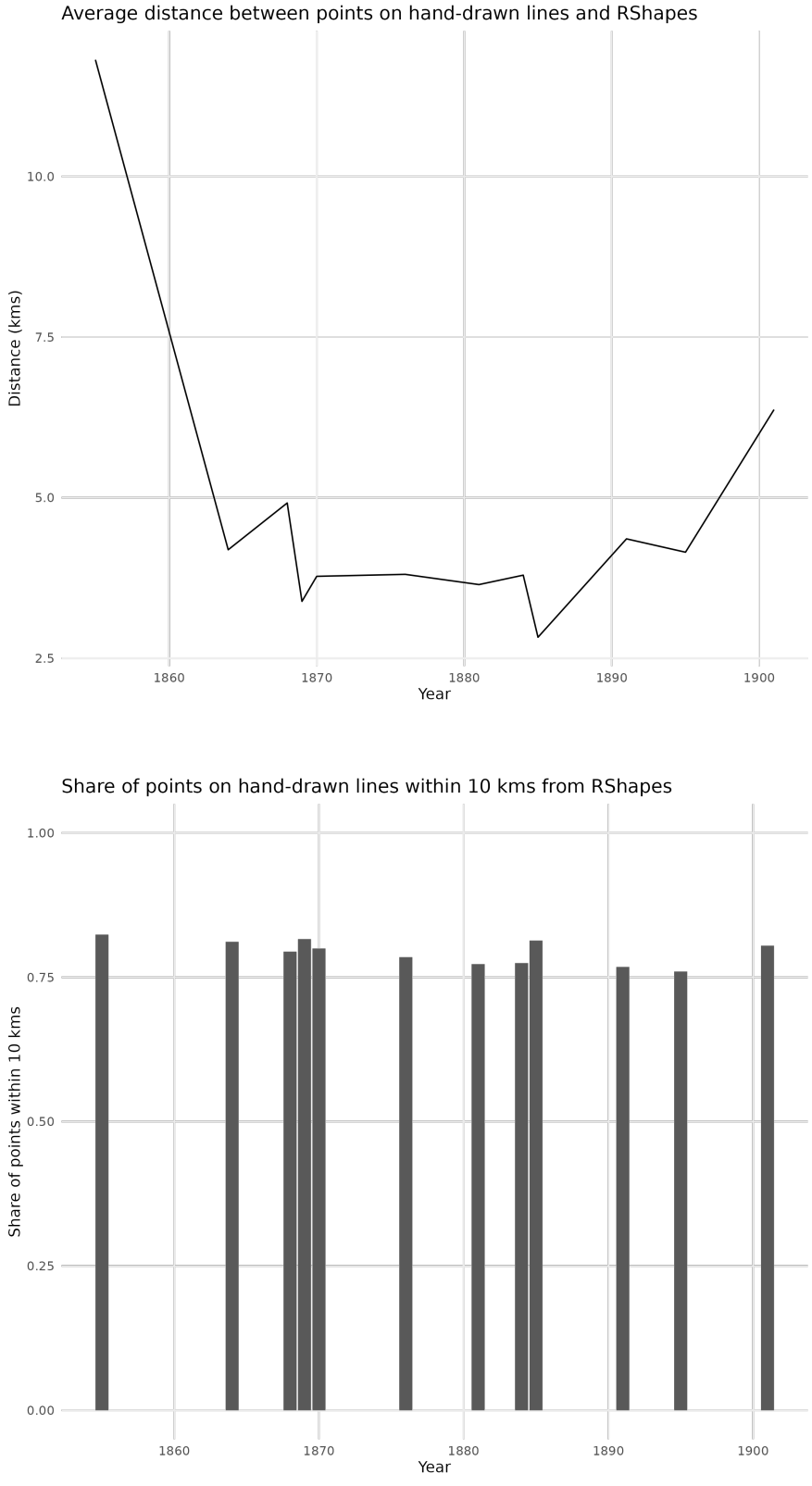


Figure A3: Similarity statistics of RShapes and Austro-Hungarian railway data.

A3 Computation of network proxies for state reach, market-based and internal connectivity

We first divide Europe in grid cells with approx. 10 km resolution, each of which is associated with a population estimate for the year 1830 (Goldewijk, Beusen, and Janssen 2010). We then build a planar graph using cell centroids as vertices and straight connecting lines to their eight queen neighbors as "footpath" edges, which we overlay and intersect with the railroad lines for each year. On the resulting graph, we can query the estimated minimum travel time between any two points in Europe for any year covered by our data. To derive the necessary edge-weights, we assume a speed of 6 km/h on "footpath" edges,[11] and 60 km/h for rail travel. The latter is close to the maximum average long-distance speeds achieved by steam-powered trains in 19th century France, Italy, and the United Kingdom.

The **state reach** proxy is calculated as a population-weighted mean of travel times between all cells in an ethnic segment and the cell that contains the respective national capital, using the 1830 population estimates.

The **market access** proxy follows Donaldson and Hornbeck (2016) and is defined as the average cell-level travel time to cities with more than 10'000 inhabitants in 1800 located in the same country. Travel times to different cities are weighted by market size (i.e., city population) and distant cities are weighted down by a trade elasticity parameter based on travel times using parameters estimated by Donaldson and Hornbeck (2016). We again aggregate cell-level market access values to ethnic segment-years by taking the population-weigthed average across all cells contained in a segment polygon. Note that market access and travel time to capital do not only vary due to local railway construction within specific segments but also as a result of population dynamics and rails built elsewhere that increase the overall connectivity within national networks. However, we chose time-invariant cell-level population weights to aggregate these measures onto the ethnic segment level to avoid potentially endogenous internal population dynamics to affect the results.

Finally, the **internal connectivity** proxy is constructed as the average travel time between any two inhabitants of a ethnic segment, again based on the 1830 population data. As longer travel times imply limited market access, weak state reach, and low levels of internal connectivity, these proxies are inverted to make sure that high values align with how we formulated our hypotheses.

A4 Data on nationalist claims

This data collection effort is inspired by the Self-Determination Movements dataset (Sambanis, Germann, and Schädel 2018), and Wimmer and Feinstein (2010). The latter code the foundation

11. Approximately the speed of horse cart travel and walking.

year of the first nationalist organization for 145 territories that were independent states in 2001. This restriction to territories that eventually became independent involves obvious selection issues which we overcome by using ethnic segments as the relevant unit of analysis.

Our coding covers all ethnic segments in historical Europe and further distinguishes the type of nationalist claims that specific nationalist organizations make. Nationalist organizations are defined as formal and non-personalistic organizations that make political claims in the name of an ethnic group. Importantly, the definition excludes cultural organizations such as national reading groups, which have been important for the expansion of literacy and national identity among rural communities, but which do not make explicit political claims (Darden 2009). We distinguish between central and peripheral nationalist claims. Central claims are either claims for minority representation in the central government or majority demands for exclusive control of the state. Peripheral claims include non-core group demands for national independence, more autonomy within the existing state, or irredentism, i.e. unification with a co-ethnic homeland abroad. For the present project, we restrict the focus to national independence and regional autonomy claims by non-core groups, as these appear as the theoretically most relevant category.

A5 Railroad simulation

We simulate railroad networks following closely the approach developed by Müller-Crepon, Hunziker, and Cederman (2021). We thus assume that states that invest in railroad infrastructure minimize the following objective function in any given year t:

$$LOSS = \frac{1}{I^2} * \sum_{j=0}^{I} \sum_{i=0}^{I} time_{j,i}, \tag{A1}$$

where $i, j \in I$ denote the inhabitants of the territory controlled by a given state who are separated by travel time $time_{i,j}$. In simple words, states aim to minimize the average travel time within their population.

To capture the pre-railroad population distribution of Europe, we turn to *estimates* in 1830 from the HYDE 3.1 data (Goldewijk, Beusen, and Janssen 2010). This estimate is derived from broad, macro level population and urbanization estimates by country (e.g., Maddison 2010), subnational census data where available, and various geographic datasets. While there is a risk that the cross-sectional differences in population are reversely affected by future railroads since part of the data is back-projected, our use of time-*invariant* population data makes it very unlikely that this would spoil our time-*variant* simulations.

Railroad investments in any state and year are constrained by the mileage of railroads we observe being built in that year in any given state territory in our Rshapes data. Because our network is much coarser and straighter than observed railroads, we deflate that budget by a

factor of 2. Each railroad line has the same quality, as we lack information on variance on that dimension.

Railroads are built by upgrading the edges of a pre-determined network of foot- and carriage paths. Given computational constraints in the repeated computation of the loss function (Eq. A1), we adjust the resolution of this baseline network to amount to .5 decimal degrees. The simulation algorithm proceeds sequentially in the following manner:

Algorithm:

1. For each state observed in t, starting at $t = 1834$, crop the Europe-wide network with all roadroads hitherto simulated to that state's territory. If the state's railroad budget for t is positive:

   (a) If no simulated railroads exist yet in the state, draw 10 seed vertices V_s with a probability proportional to their population. Sample one incident edge per vertex V_s and upgrade it to become a railroad "seed edge" and part of the collection of built lines E_b. Subtract length of built lines from budget.

   (b) Select all neighboring edges of E_b, evaluate their impact on *LOSS* and keep 10 most promising edges as E_p.

   (c) Upgrade edge $e \in E_p$ that minimizes *LOSS*. Select neighboring edges of e that have not yet been upgraded and add to E_p. Update $B_q = B_q - length_e$.

   (d) Repeat step (c), and, in every 10^{th} round, step (b), until budget B_q for a given state in year t is spent.

2. Move to the next year, t = t+1, until arriving in 1922, the last year covered by our railroad data.

A6 Additional plots and tables

Interaction Plots

Figure A4: Marginal Effect Plot & Binning Estimates: Ling. Distance

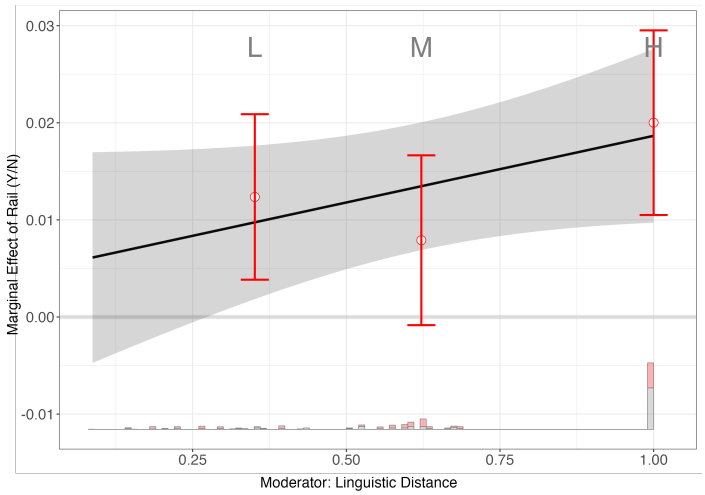


Figure A5: Marginal Effect Plot & Binning Estimates: Share Core Group

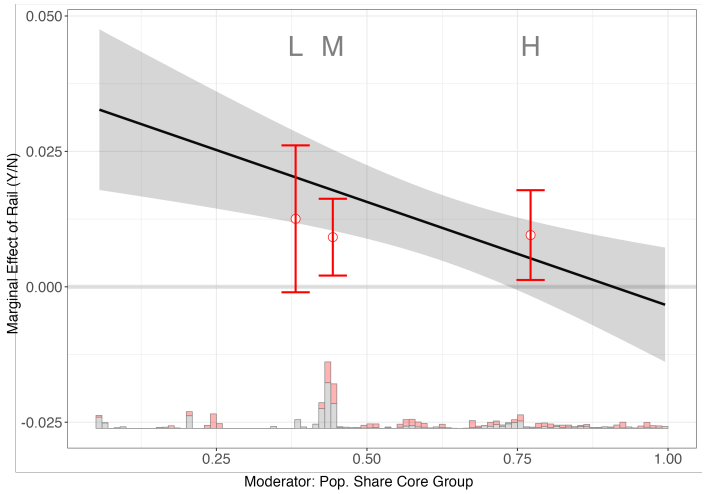


Figure A6: Marginal Effect Plot & Binning Estimates: Segment Population

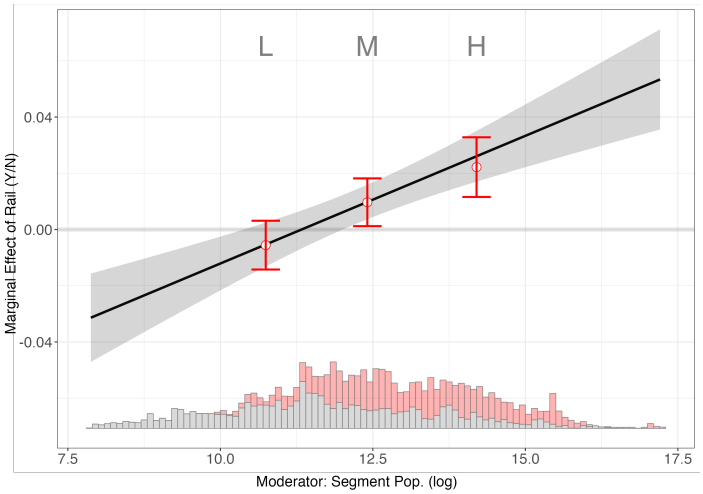


Figure A7: Marginal Effect Plot & Binning Estimates: Per capita GDP

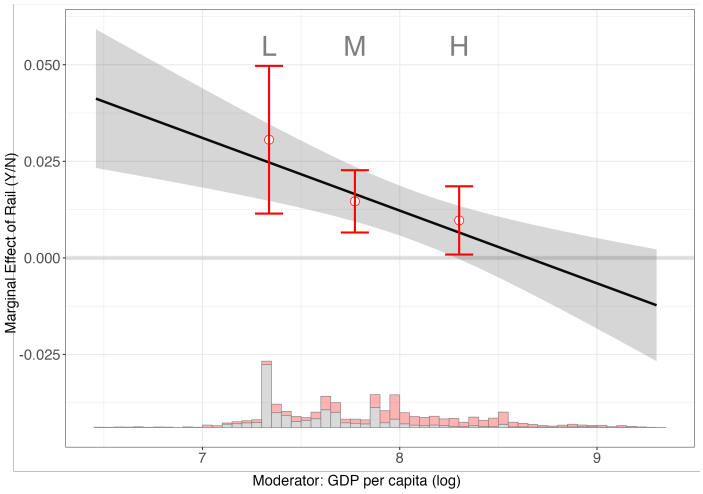


Figure A8: Marginal Effect Plot & Binning Estimates: Fiscal capacity

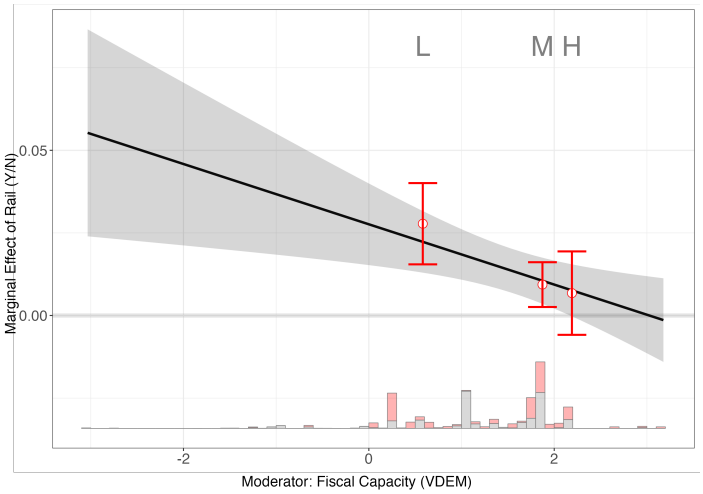
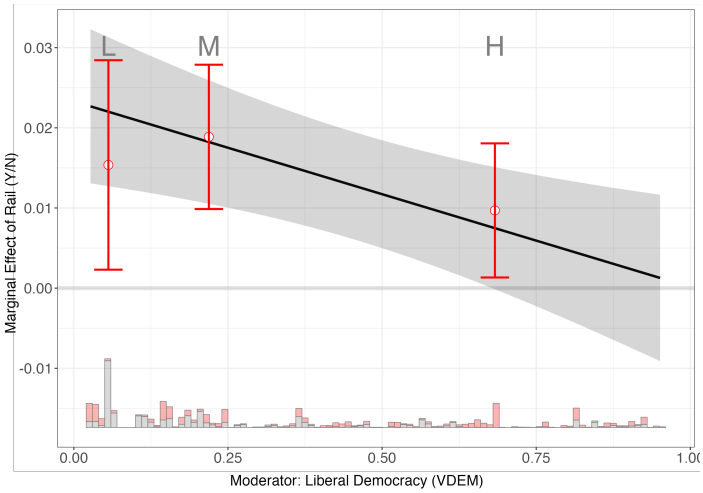


Figure A9: Marginal Effect Plot & Binning Estimates: Liberal democracy



Robustness tests

Table A1: DiD Models: Dropping Cases with Past Separatism

	100 × Secession, Terr. CW or Claim			
	(1)	(2)	(3)	(4)
Rails (Y/N)	1.247**	0.965**	0.893**	0.924**
	(0.386)	(0.366)	(0.302)	(0.333)
Segment FE	Yes	Yes	Yes	Yes
Year FE	Yes	No	Yes	No
Country-Year FE	No	Yes	No	Yes
Estimator	TWFE	TWFE	2S-DiD	2S-DiD
Mean DV	1.037	1.037	0.941	0.9
Observations	8679	8679	7650	6667

Notes: The unit of analysis is the ethnic segment year. State-leading segments, segments smaller than 2000 sqkm, and those with past secessionsit civil war and claims for independence or autonomy dropped. Segment clustered standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A2: DiD Models: Dropping Never-Treated Units

	100 × Secession, Terr. CW or Claim			
	(1)	(2)	(3)	(4)
Rails (Y/N)	1.664***	0.702*	4.616***	4.477***
	(0.430)	(0.340)	(1.055)	(0.938)
Segment FE	Yes	Yes	Yes	Yes
Year FE	Yes	No	Yes	No
Country-Year FE	No	Yes	No	Yes
Estimator	TWFE	TWFE	2S-DiD	2S-DiD
Mean DV	1.215	1.215	1.168	1.217
Observations	11 114	11 114	9759	7479

Notes: The unit of analysis is the ethnic segment year. State-leading segments and segments smaller than 2000 sqkm dropped. Segment clustered standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A3: Network Structure (Country-Year Fixed Effects)

	100 × Secession, Terr. CW or Claim			
	(1)	(2)	(3)	(4)
National Market Access	−0.091			−0.021
	(0.104)			(0.102)
State Reach		−0.012***		−0.013***
		(0.003)		(0.003)
Internal Connectivity			0.014+	0.012+
			(0.007)	(0.007)
Segment FE	Yes	Yes	Yes	Yes
Country-Year FE	Yes	Yes	Yes	Yes
Mean DV	1.131	1.115	1.115	1.114
Observations	12 643	13 007	13 007	11 652

Notes: The unit of analysis is the ethnic segment year. State-leading segments and segments smaller than 2000 sqkm dropped. Segment clustered standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A4: Instrumenting Railroads: Country-Year FE

	Rails (Y/N)	100 × Separatism		
	First Stage	OLS	Reduced Form	Second Stage
Rails (Y/N, simulated)	0.279***		0.904**	
	(0.064)		(0.318)	
Rails (Y/N)		1.111**		
		(0.364)		
Rails (Y/N, instrumented)				3.242*
				(1.259)
Segment FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
First Stage F	18.781			18.781
Mean DV	0.512	1.115	1.115	1.115
Observations	13 007	13 007	13 007	13 007

Notes: The unit of analysis is the ethnic segment year. State-leading segments and segments smaller than 2000 sqkm dropped. Segment clustered standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Disaggregating the separatism outcome

Table A5: Railroads and Secession (1816-1945)

	100 × Secession			
	(1)	(2)	(3)	(4)
Rails (Y/N)	0.023	−0.051	0.296***	0.304***
	(0.079)	(0.091)	(0.071)	(0.090)
Segment FE	Yes	Yes	Yes	Yes
Year FE	Yes	No	Yes	No
Country-Year FE	No	Yes	No	Yes
Estimator	TWFE	TWFE	2S-DiD	2S-DiD
Mean DV	0.131	0.131	0.145	0.122
Observations	13 007	13 007	11 711	9818

Notes: The unit of analysis is the ethnic segment year. State-leading segments and segments smaller than 2000 sqkm dropped. Segment clustered standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A6: Railroads and Separatist Conflict (1816-1945)

	100 × Terr. CW			
	(1)	(2)	(3)	(4)
Rails (Y/N)	0.832***	0.517**	1.168***	0.850**
	(0.216)	(0.190)	(0.340)	(0.261)
Segment FE	Yes	Yes	Yes	Yes
Year FE	Yes	No	Yes	No
Country-Year FE	No	Yes	No	Yes
Estimator	TWFE	TWFE	2S-DiD	2S-DiD
Mean DV	0.438	0.438	0.478	0.519
Observations	13 007	13 007	11 711	9818

Notes: The unit of analysis is the ethnic segment year. State-leading segments and segments smaller than 2000 sqkm dropped. Segment clustered standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A7: Railroads and Separatist Claims (1816-1945)

	100 × Independence or Autonomy Claim			
	(1)	(2)	(3)	(4)
Rails (Y/N)	0.625*	0.611+	0.659*	0.583+
	(0.282)	(0.327)	(0.296)	(0.345)
Segment FE	Yes	Yes	Yes	Yes
Year FE	Yes	No	Yes	No
Country-Year FE	No	Yes	No	Yes
Estimator	TWFE	TWFE	2S-DiD	2S-DiD
Mean DV	0.569	0.569	0.478	0.458
Observations	13 007	13 007	11 711	9818

Notes: The unit of analysis is the ethnic segment year. State-leading segments and segments smaller than 2000 sqkm dropped. Segment clustered standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

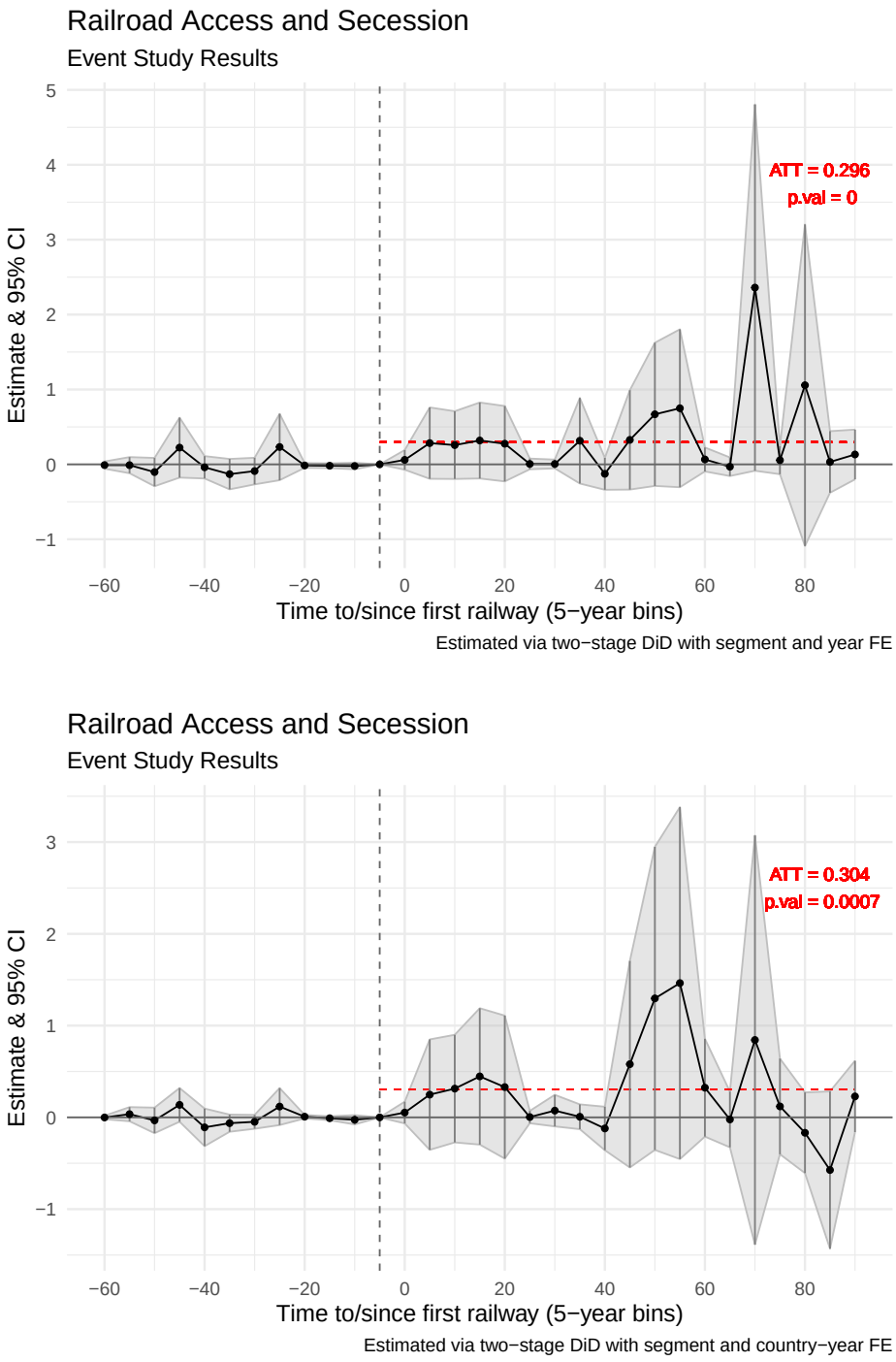
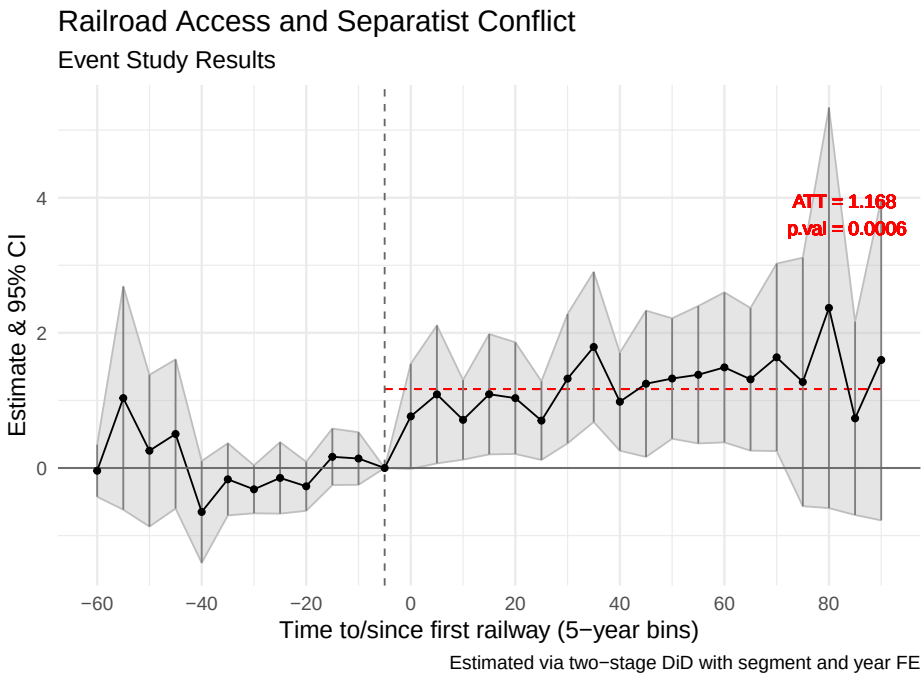


Figure A10: Event study plots
(ATT estimates based on Columns 3 and 4 in Table A5)



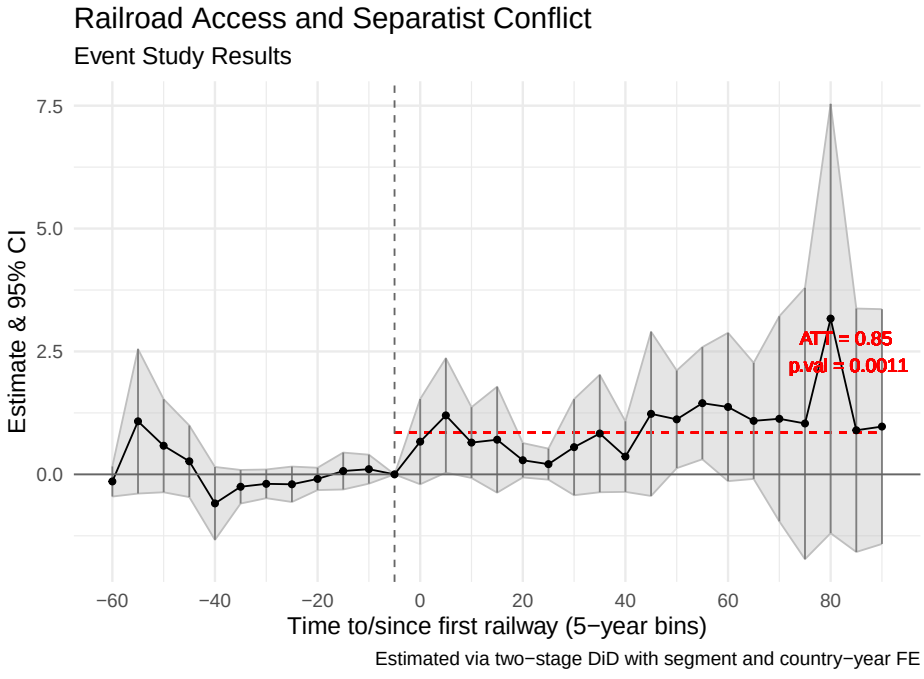
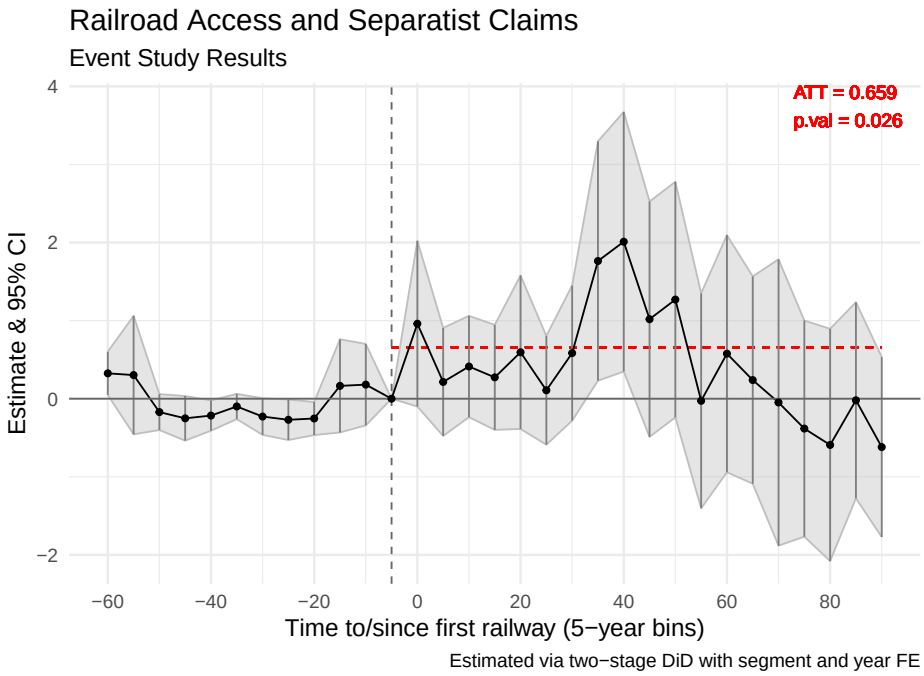


Figure A11: Event study plots
(ATT estimates based on Columns 3 and 4 in Table A6)



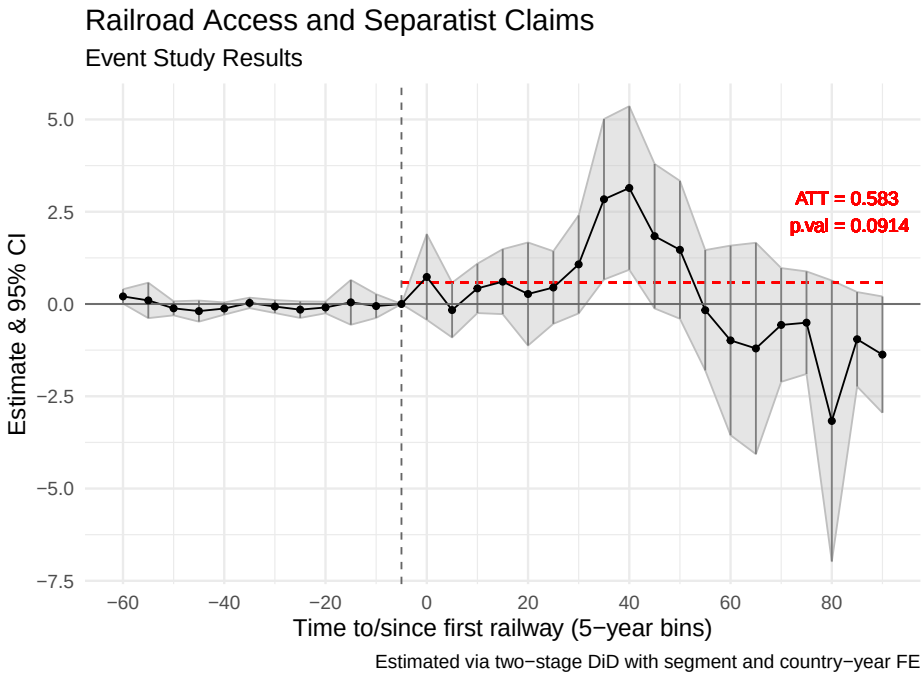


Figure A12: Event study plots
(ATT estimates based on Columns 3 and 4 in Table A7)

References

Darden, Keith. 2009. "Resisting Occupation: Lessons from a Natural Experiment in Carpathian Ukraine." Unpusblished manuscript, November.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American economic growth: A "market access" approach." *The Quarterly Journal of Economics* 131 (2): 799–858.

Goldewijk, Kees Klein, Arthur Beusen, and Peter Janssen. 2010. "Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1." *The Holocene* 2010 (1): 1–9.

Lewis, M. Paul, ed. 2009. *Ethnologue: Languages of the world.* Vol. 16. SIL International Dallas, TX.

Maddison, Angus. 2010. "Maddison Historical Statistics of the World Economy." *Groningen, Países Bajos: University of Groningen.*

Müller-Crepon, Carl, Philipp Hunziker, and Lars-Erik Cederman. 2021. "Roads to rule, roads to rebel: relational state capacity and conflict in Africa." *Journal of Conflict Resolution* 65 (2-3): 563–590.

Sambanis, Nicholas, Micha Germann, and Andreas Schädel. 2018. "SDM: A New Dataset on Self-Determination Movements with an Application to the Reputational Theory of Conflict." *Journal of Conflict Resolution* 62 (3): 656–686.

Wimmer, Andreas, and Yuval Feinstein. 2010. "The Rise of the Nation-State across the World, 1816 to 2001." *American Sociological Review* 75 (5): 764–790.