# 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** *Princeton University*
roberto.valli@princeton.edu

**Lars-Erik Cederman** *ETH Zurich*
lcederman@ethz.ch

**Luc Girardin** *ETH Zurich*
luc.girardin@icr.gess.ethz.ch

September 7, 2024

**Abstract.** This paper uses the gradual expansion of the European railway network to investigate how this key technological driver of modernization affected ethnic separatism between 1816 and 1945. 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. These effects concentrate in countries with small core groups, weak state capacity, and low levels of economic development as well as in large ethnic minority regions. Exploring causal mechanisms, we show how railway networks can facilitate mobilization by increasing the internal connectivity of ethnic regions and hamper it by boosting state reach. Overall, our findings call for a more nuanced understanding of the effects of European modernization on nation building.

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, drove 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 promoted national integration and identity formation (Deutsch 1953; Gellner 1983; Anderson 1983). Simultaneously, expanding transportation networks contributed to separatist mobilization of culturally distinct peripheral groups (Hechter 2000; Breuilly 1982; Huntington 1968).

In this paper, we investigate how the expanding European railway network contributed to nationalist mobilization that either united or divided states. Our theoretical argument builds on and extends the existing literature on modernization, nationalism, and separatism. We specify three mechanisms through which railroads may affect competition and bargaining between the central state and ethnically distinct peripheral regions. 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 separatism. Since the integrative processes of cultural assimilation, state-led nation-building, and economic modernization tend to unfold more slowly 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, state reach, and local mobilization capacity. In addition, we study how ethnic demography, economic development, and political institutions affect whether railroad construction caused national integration or disintegration.

We test these arguments by combining newly collected geo-spatial data on the expanding European railway network (1834–1922) with measures of independence claims, secessionist civil wars, and successful secession (1816–1945). We link these data to yearly observations of ethnolinguistic group segments derived by intersecting historical maps of ethnic settlements 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. This effect materializes immediately and dissipates over time without turning negative. In addition to observing parallel pretreatment trends, an instrumental variable approach based on simulated railroad networks bolsters the robustness and causal interpretation of our findings. Second, our analysis of heterogeneous effects shows that separatist responses to railway access complicate top-down nation-building in states with low levels of economic development and state capacity while providing motivations and opportunities for national independence campaigns, in particular among large minorities. Third, a disaggregated analysis of mechanisms underlying the effect of railway access suggests that improvements in state reach reduce separatism, internal connectivity increases the risk, with market access exerting little effect.

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 (Weber 1976) and disintegration and separatist nationalism in Eastern Europe (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 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. Yet recent empirical contributions link railroads to the diffusion of opposition movements (Brooke and Ketchley 2018; García-Jimeno, Iglesias, and Yildirim 2022; Melander 2021) and resistance to the state (Pruett 2023).[1]

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

## Modernization and nationalism in the literature

The introduction of steam-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). The relevant arguments fall into two main camps depending on whether they stress national integration or separatism.

The former school expects cultural homogenization and increasing identification with the state-leading nation (see e.g. Robinson 2014; Eifert, Miguel, and Posner 2010). Political accounts highlight the modern state as the key agent of change (Hobsbawm 1990). On this view, states devise and implement nation-building programs to respond to both international and domestic threats (Hintze 1975; Tilly 1994; Posen 1993). A complementary perspective views the develop-

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).

ment of industrial economies as the main integrating force. In Gellner's (1983) seminal account, the transition from agrarian to industrial modes of production requires standardized languages (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 industrial drivers of nationalism.

Despite their integrationist thrust, modernist accounts also shed light on national disintegration. Adopting a political perspective, Breuilly (1982) and Hechter (2000) expect the shift from indirect to direct rule to trigger reactive mobilization, especially where peripheral elites enjoyed autonomy prior to state centralization. Similarly, Deutsch (1953) notes that wherever social mobilization outpaced assimilation, nationalist conflict became more likely. Gellner (1983, 1964) expects the combination of pre-existing cultural difference and uneven development to trigger separatism.

Complementing the theoretical classics, several empirical studies analyze, albeit selectively, the link between modernization and nationalist mobilization. Perhaps most famously, Eugene Weber (1976) traces French national identity formation in the 19th century, highlighting industrialization, expanding transportation and communication networks, and state policies as integrating forces. Despite his brilliance, however, Weber (1976) remains a historian of France, a country that enjoyed particularly successful nation-building compared to most other European countries.

More recently, cross-country studies show that state-led nation-building efforts, in particular education reforms, become more likely when rulers faced international (Aghion et al. 2019) or domestic threats (Paglayan 2022; Alesina, Giuliano, and Reich 2021). While these studies explain the strategic timing of nation-building policies, the mere adoption of such efforts does not guarantee their success.

Micro-level quantitative work within single countries illustrates how specific educational, linguistic, and religious state-building efforts succeeded or backfired in 19th century and contemporary France (Balcells 2013; Abdelgadir and Fouka 2020), Prussia (Cinnirella and Schueler 2018), colonial Mexico (Garfias and Sellars 2021), 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 disintegration but say less about cross-country variation.

In one of the very few comparative studies, Wimmer and Feinstein (2010) focus on nation-state creation in a global sample of 145 territories corresponding to independent states in 2001 back-projected until 1816. Using railway density as a modernization proxy, they find no effect on the transition to nation-states in pre-national or newly independent states. Despite this pioneering effort, their over-aggregated research design suffers from hindsight bias due to the backward-projected sampling based on contemporary state units which were shaped along ethnic lines as a results of nationalist border change (Müller-Crepon, Schvitz, and Cederman 2023).

In sum, then, the link between modernization and national integration and disintegration remains contested. First, scholars disagree about whether modernization spurs nationalism for or against the state and what mechanisms account for the link between modernization processes and nationalist mobilization. Second, the existing literature provides little theoretical or empirical guidance as regards the contextual factors that produce state-building or counter-state nationalism in specific cases. Third, while the classic contributions offer little systematic evidence for their claims, the recent micro-level studies convincingly validate parts of the classical theories in selected countries, but offer no comparative outlook.

The present paper addresses these three gaps in the existing literature. First, we analyze railway construction to assess whether this crucial technological driver of modernization has systematically produced national integration or disintegration. Second, we study the effect of causal mechanisms and contextual factors that contribute to national integration or counter-state nationalism. Third, our Europe-wide data are spatially disaggregated at the subnational level, thus allowing us to integrate the literatures relying on cross-country comparisons and micro-level analysis of individual cases.

## Railways and nationalist mobilization

As our discussion of existing research shows, railway expansion and the associated 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 and 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 oppositional mobilization and spur separatist responses to national integration.

Our theoretical framework draws on the literature reviewed above to explain how, and under what conditions, railroad construction united or divided Europe's multi-ethnic states. We introduce 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, typically demographically smaller, and more peripherally located than their state-leading counterparts (Mylonas 2012). Practically all states in Europe contained such minority segments. Before industrialization, central governments typically ruled non-core groups indirectly by 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).[2]

The situation changed when industrialization, direct forms of rule, and nationalist ideologies swept across Europe in the 19th century. Separatist mobilization occurred wherever elites of non-core group managed to rally their followers against the state. Benefiting from agrarian

2. Historians refer to these units as "composite monarchies" (Elliott 1992). Even metropolitan France, arguably the most centralized and cohesive state in the early 19th century, had imperialist traits (Weber 1976).

economies and indirect rule, some leaders belonged to old elites, whose status was threatened by local industrialization or state centralization (Hechter 2000; Garfias and Sellars 2021). Other leaders made up "new elites", ranging from bourgeois liberals and democratic reformers to ethnonationalists (Gellner 1983; Hutchinson 1987).

For these new and old 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 unaligned (Breuilly 1982). Third, once ideologies of national self-rule took root, bravely resisting domination by a culturally foreign elite allowed them to mobilize local populations more effectively than alternative opposition frames (Balcells, Daniels, and Kuo 2023; Gellner 1983). Lastly, separatist mobilization raised the prospects of securing support from nationalizing Europe's great powers, which became increasingly receptive to ideals of national self-determination (Breuilly 1982).

Taking separatism as the main outcome under investigation circumvents the challenge of 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 whether crucial, necessary conditions for integration are absent or, in other words, zoom in on clear failures of nation building. Wherever a culturally distinct region breaks away from a state or mobilizes the local population in an attempt to do so, nation building has evidently failed.[3]

3. Yet the absence of separatism is clearly not a sufficient condition for national integration (Connor 1972).

Among the forms that separatist mobilization can take, we consider the formation of organizations claiming autonomy or independence for an ethnic group, as well as attempted or successful secessions. While some separatist movements never went beyond making nationalist claims, such as the demands for autonomy by Spanish Galicians in the 1930s (Garcia-Alvarez 1998), other movements escalated violently. In the Ottoman Empire, for instance, Bulgarians and Romanians successfully gained independence through the 1878 Treaty of Berlin. In both cases, initial independence claims were followed by secessionist civil war in the 1870s (Minahan 2001; Goina 2005, 137).

## Motivations Driving Railroad 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 unrepresentative in this respect. 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 country's periphery (Weber 1976). The centralizing logic was also present in Sweden (Cermeño, Enflo, and Lindvall 2022), Belgium, and with major delay, Spain (Alvarez-Palau, Díez-Minguela, and Martí-Henneberg 2021). In unifying Germany and Italy, railroad construction contributed to 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).

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. Besides limited 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 also 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. Yet the belated drive for nation-building and Russification gave railroads a prominent role as cultural homogenizers. As these different motivations of railroad construction may potentially be related to past or future separatism, the empirical analyses below include different strategies account for endogenous railroad expansion.

## 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 support separatist movements. This choice depends on the expected costs, benefits, and chances for success of state-led nation-building and national independence campaigns. Railway construction in the periphery may thus affect the emergence of separatist movements if it shifts these costs, benefits, and success probabilities as perceived by local populations. Here 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 before deriving contextual factors that may tilt the balance in favor of integration or disintegration.

The three 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). The following paragraphs lay out how growing railroad networks, through these three mechanisms, affect the costs and benefits, as well as the likelihood of success of separatist mobilization.

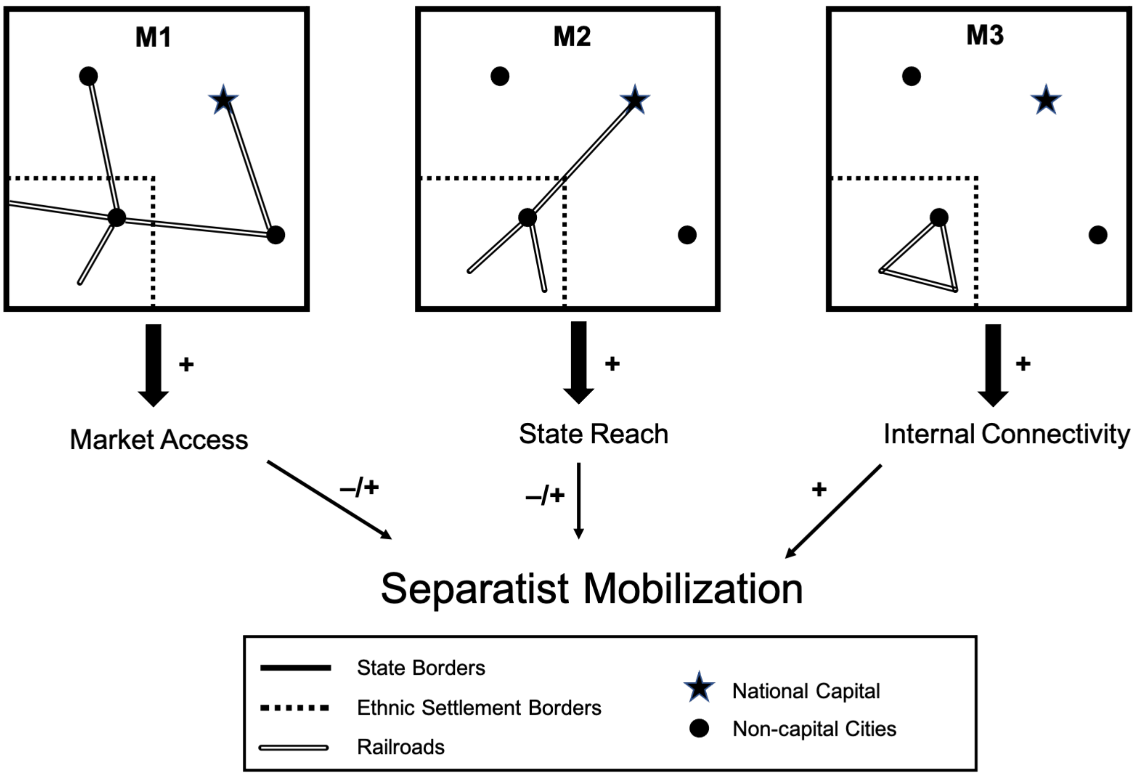



Figure 1: How railroad construction may matter

**M1: Market access and social communication.** First and foremost, railroads affect local populations through economic integration and social communication. Improved connectivity to the entirety of a country's territory, and especially to major cities, increases the costs of secession by making economic independence less attractive. It instead 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 development is inextricably linked with railway construction.[4] Moving goods and people across large distances enabled the formation of integrated market economies and labor migration from agrarian towns to industrializing cities (Rostow 1960; Fishlow 1965; Weber 1976). 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 separatist elites' attempts to cut them off from these emerging opportunities (Hierro and Queralt 2021).

Railways accelerated the expansion of communication networks, brought previously isolated rural residents in contact with urban dwellers and each other, thus creating the bottom-up incentives and pressures for cultural homogeneization described by Gellner (1983) and Deutsch (1953). Weber (1976, ch. 12) describes road and railway networks as technological precondition for "radical cultural change" in nationalizing France (see also Segal 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. Examples include minorities in integrating, western states, such as the Catalans in France and the German and Frisians in the Netherlands.

However, cultural difference may become more salient where members of distinct ethnic groups compete for inherently scarce modernization benefits (Bates 1983). Similarly, Gellner

4. Of course, improvements of other means of communication also contributed to this process, such as road and canal construction (see e.g. Fogel 1964).

(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 could increase support for separatist movements. While such a "backlash"-effect is less prominently discussed in the literature, railroad expansion can, in principle, also increase peripheral populations' motivations thus reducing the costs of elite-led separatist mobilization. This dynamic might be particularly acute in geographically isolated segments that experience large increases in domestic market access due to railway construction, such as the Finns who gained independence from the Russian Empire in 1917.

**M2: State reach and direct rule.** A second and plausibly equally important mechanism links railroads to the central state's ability to reach, govern, and transform local populations in top-down fashion. Providing public goods and engaging in ambitious state- and nation-building policies would have been inconceivable without railroads (Wimmer 2018). Modern transportation infrastructure is part of 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 reduces 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). The prospect of state-repression increases the costs of separatist mobilization and lowers the changes of separatist success. 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 as intended, local populations should have less motives and opportunities to support separatism. The Austrian-Hungarians successful expansion of mass education to the Ukrainian parts of the Habsburg Empire fits this pattern (see e.g. Darden 2009), as do the French efforts to assimilate its periphery, including the Basques.

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), thus nurturing popular and elite-level support for secession and facilitating separatist mobilization. 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). The Russian Empire's expansion of rail connections to the Polish lands facilitated separatist mobilization including among railroad workers (Schenk 2011). The Tanzimat reforms in the Ottoman Empire were met by Serb resistance in 1878 and 1910 (Hechter 2000; Malesevic 2012).

**M3: Internal connectivity and social mobilization.** Third, railroads can facilitate the coordination and collective action of peripheral opposition movements, thus lowering the costs of separatist mobilization. Mechanism M3 in Figure 1 shows how local rails within a culturally distinct subregion 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). Similarly, denser peripheral road networks come with higher levels of organized violence against the state in Africa Müller-Crepon, Hunziker, and Cederman (2021). 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, often unintentionally, railroad construction may thus increase the opportunities for separatist mobilization and, via internal communications and exchange, promote identification with separatist movements. Reactive mobilization occurred in groups that were traversed by the state's main railroads network, such as the Ukrainians and Belorussians in Tsarist Russia and the Bulgarians in the Ottoman Empire. Even some industrializing segments in Western Europe, such as the Catalans in Spain, benefited from increasing levels of internal connectivity and managed to resist the assimilationist and integrationist advances of the central state.

**Deriving testable hypotheses.** 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 hand, both market integration (M1) and state penetration (M2) 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 typically evolves over a long time period. In France, this process lasted for a full century following the French Revolution (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). In contrast, backlash against market integration and state-building often occurs immediately upon their arrival.

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. [5] 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.*

5. Although depicting an overall slow-moving process of assimilation into French national identity, Weber (1976, 205–207) stresses highlights the first arrival of a rail connection in a locality as a mind-opening, perhaps even revolutionary event that abruptly pushed rural areas out of their pre-modern slumber. In ethnically distinct areas, this shock often provided a trigger for counter-state mobilization.

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.

**Conditional hypotheses.** 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 2** *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.*

**Network structure and specific causal mechanisms.** Finally, we move beyond the short-term effects of 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 potentially 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 3** *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).*

## 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.[6]

**Ethnic settlement data.** Information on historical ethnic settlements comes from the newly compiled Historical Ethnic Geography (HEG) dataset which is based on a selection of 73 historical maps (for details, see online Appendix A1). Practically all ethnic categories appearing on our maps refer to linguistic rather than religious or regional ethnic identity markers, thus reflecting a well-known characteristic of European nationalism (Barbour and Carmichael 2000). We standardize all groups depicted on all maps with the help of the Ethnologue language tree (Lewis 2009) and construct a time-invariant master list. Finally, we draw on all maps belonging to a specific group-time period combination to construct a best-guess settlement polygon.

**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 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*) from 1816

6. Replication materials can be found on the APSR Dataverse (Valli 2024).

to 1945. 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 the largest ethnic segment that contains the capital, subject to manual inspection and correction.

**Main independent variable: Railway access.** Segments' access to railway networks serves as a geographically and temporally disaggregated proxy for the uneven spread of modernization. Geographic data on the expanding European railway network comes from train.eryx.net, a website built by French train enthusiasts Bernard and Raymond Cima. They provide construction dates and map representations of all known railway segments covering almost all of geographic Europe, with the notable exception of England and Wales, which we exclude from the analysis. We georeference their yearly online map tiles and digitize all line features to construct a geospatial dataset of European rails from the first railway built in 1834 to 1922.[7] Figure 2 plots our railroad data. Appendix A1 validates the railroad data's precision against 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 line datasets of the European railway network. All segments intersected by a line feature are assumed to be connected. To operationalize mechanisms M1 to M3, we use the network structure to compute continuous proxies for segments' connectivity to national economic markets, state reach, and internal connectivity (see Appendix A4 for details).

7. Our yearly resolution improves upon the decade-level coding of railroad networks in Martí-Henneberg (2021) and Alvarez-Palau, Díez-Minguela, and Martí-Henneberg (2021).

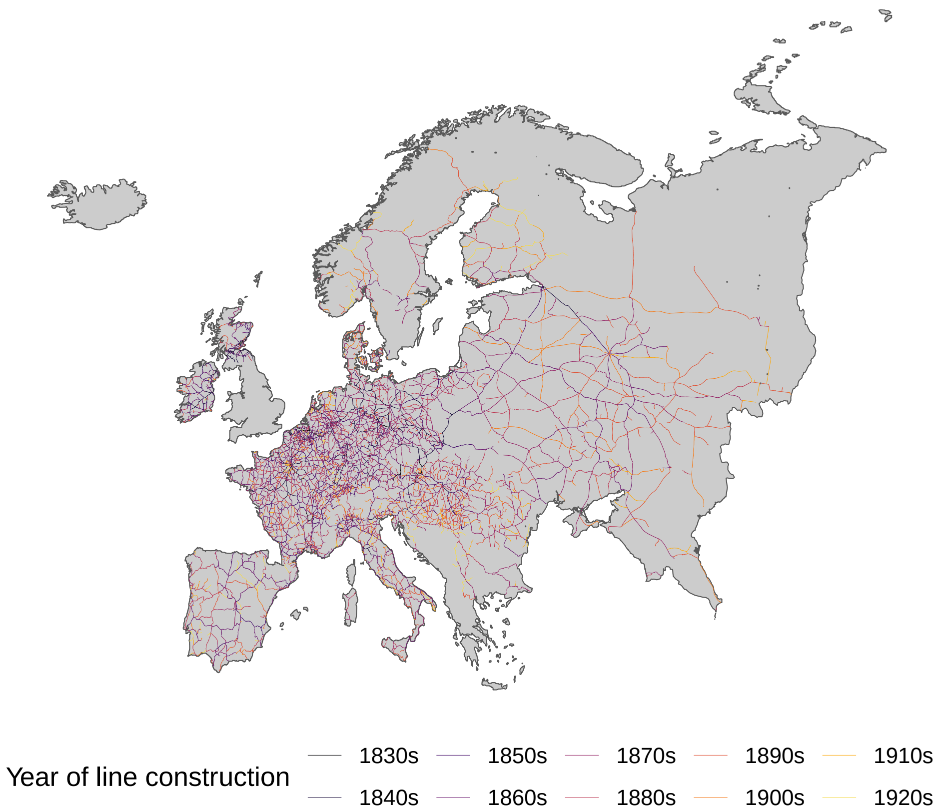



Figure 2: Geographic data on yearly railway construction (digitized from train.eryx.net).

**Outcomes: Attempted and/or successful secession.** As described below, our main outcome variable captures violent and peaceful mobilization for separatism by combining onsets of separatist conflict, successful secessions, and political claims for national independence or regional autonomy (see Appendix A3.1).

First, we code a dummy of *ethno-territorial civil war onset* at the ethnic segment-year level. For the period 1816–1945,[8] we identify all unique civil wars listed in the datasets provided by Gleditsch (2004) and Sarkees and Wayman (2010) that were fought in the name of a specific ethnic group, focusing on ethnic claims and recruitment.

8. We include years beyond the coverage of our railroad data to study the long-run effects of railway expansion. Appendix A8 shows stronger results for the period 1915-1922.

We combine the territorial conflict measure with a binary indicator of *successful secession* as an additional signal of national disintegration.[9] The secession dummy is coded one for all non-core ethnic segments that become core group segments in newly independent states in year $t + 1$.

Lastly, we add a new measure coding *nationalist claims* to code the first claim for full national independence or regional autonomy within given state structures made by a nationalist organization at the level of ethnic segment years (see Appendix A5). In combination, the disintegration measure takes on the value of 1 if a segment experiences a secessionist conflict onset, claim, or secedes in a given year and 0 otherwise. Table A1 in the Appendix presents descriptive statistics of all main dependent and independent variables.

# Analyses and results

This section summarizes our main specification and results, followed by a set of robustness checks. We then test our conditional hypotheses and present results on disaggregated mechanisms tests.

## Main Specification 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

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

years or country-years – the latter control for the potential of regionally concentrated diffusion of secessionism 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 that counterfactual trends are parallel, 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 effects evolving dynamically after the first treatment onset (e.g. Goodman-Bacon 2021; Callaway and Sant'Anna 2021; Roth et al. 2023; Chaisemartin and D'Haultfœuille 2022). Therefore, we also implement two-stage estimators recently proposed by Gardner (2021) and Liu, Wang, and Xu (2024), which are specifically suited to multi-cohort DiDs with staggered treatment adoption. By imputing counterfactual outcomes for treated units based on a first-stage regression, the 2S-DiD approach alleviates most of the weighting and comparison problems of conventional TWFE models. Appendix A7 describes our choice of estimators in detail and shows robustness to alternative DiD specifications (Liu, Wang, and Xu 2024).

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 secession increases by 1.49 percentage points after the first railway arrives. This effect is substantively large and amounts to a more than two-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). Both specifications yield substantively larger estimates than their TWFE-based counterparts in

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) | (0.446) |
| 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. All models control for the number of past conflicts and peace years indicators. Segment clustered standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Columns 1 and 2. The difference in magnitude can be explained by the mechanical downward biases that TWFE models create in staggered treatment settings when temporal effect heterogeneity exists (e.g. Goodman-Bacon 2021, p. 261). Model 3 suggests an effect of railroads of more than 2 percentage points, equivalent to a 195% increase from the sample mean. The effects 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 its plausibility.

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, five-year long

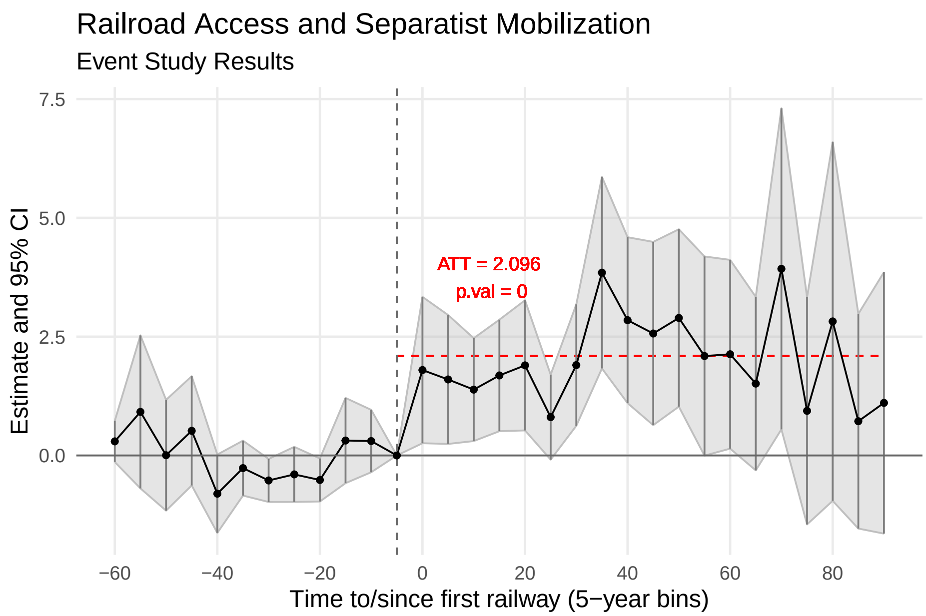



Figure 3: Event study plot
(ATT estimates based on Column 3 in Table 1)

time-to-treatment bins.[10] The first five-year bin before treatment onset is omitted and serves as the baseline category. We obtain similar results from an event study model using country-year instead of year fixed effects (see Figure A6 in the Appendix). Both plots reveal mostly 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 Figure 3 are negative and significant at the 5% level, which is not the case when using country-year fixed effects. The parallel trends prior to treatment make the identifying assumptions of our empirical strategy more plausible and should reduce concerns about endogenous railway building in response to separatist mobilization. The post-treatment dummies indicate an immediate increase in conflict risk after the first railway is built. The estimated treatment effects grow even larger

10. For each segment that eventually receives railways, we code (i) whether a 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.

approx. 35 years after the first railway and, if anything, diminish from the $50^{\text{th}}$ post-treatment year onward (especially in the specification with country-year FE). These results clearly support Hypotheses 1 and cast 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

**Instrumental variable approach.** An instrumental variables (IV) strategy based on simulated railways addresses remaining potentials for reverse causality and omitted variable bias, by which security considerations or other proximate causes of conflict motivate railway extensions. We simulate the evolution of railway networks by heuristically placing railroads for each country-year such that they maximize the connectedness of a state's population (see Appendix A6). The simulated development of the European railroad network is thus only determined by the yearly mileage built in each state, their borders, as well as the *time-invariant* population distribution as estimated for 1830 (Goldewijk, Beusen, and Janssen 2010), thus excluding potentially biasing military, demographic, or economic causes of railroad construction.

We use the presence of a simulated railroad in a segment as an instrument for observed railway access in a TWFE estimation strategy. The exclusion restriction assumes that the instrument affects separatism only through observed railroads and is not systematically affected by unobserved causes of conflict. Our segment fixed effects account for potential time-invariant omitted variables and year fixed effects capture temporal fluctuation in railroad expansion. We additionally show robustness to country-year fixed effects which account for state-specific railroad investments and border changes.

Column 1 in Table 2 shows that our instrument is strongly predictive of actual railway construction in ethnic segments (F-stat of 39). Column 2 replicates our TWFE baseline to facilitate comparing naive to IV estimates. Columns 3 shows the reduced form regression of separatism

Table 2: 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. All models control for the number of past conflicts and peace years indicators. Segment clustered standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

on the instrument, whereas Column 4 shows the second-stage estimate of instrumented rail access. Both coefficients are positive and statistically significant, yet less precisely estimated than the baseline TWFE effect. The second stage yields an estimate larger then the TWFE but similar in size to the 2S-DiD estimate (Table 1, Model 3). Replacing year with country-year fixed effects leads to stronger results (Appendix Table A3). These findings increase our confidence that the estimated effects are not merely reflecting reverse causation resulting from strategic railway construction or biases from temporally varying omitted variables.

**Sample definitions.** As an alternative to controlling for past conflict in our baseline models, we run a robustness check that drops all ethnic segment-years as soon as they experience a secessionist civil war or nationalist claims. The results are summarized in Table A4 and show substantively smaller, yet positive and significant, treatment effects. These estimates also treat all separatism outcomes equally by censoring observations after the first onset of separatism. Therefore the models can be interpreted as the effect of railways on the risk of separatism given no previous separatist effort. 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 1922 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 A5 shows similar or, when using the two-stage DiD estimator, significantly larger treatment effects. Finally, we replicate our main findings by censoring the sample in 1922, year in which our railway data stops. Results in Tables A6–A7 and Figure A7 are remarkably robust, show estimates that are of the same or larger size than in the main specification.

**Outcome disaggregation.** We furthermore disaggregate the outcome variable and report separate regressions for successful secessions, secessionist civil wars, and national independence or autonomy claims. The results in Appendix Tables A9–A10 suggest that our baseline findings are mainly driven by territorial civil wars and nationalist claims. That said, the estimated effects on the most extreme (and rare) outcome of successful secession are positive and reach significance when estimated as two-stage DiD but substantively small and insignificant in the TWFE setup.

**Including irredentism.** The combined outcome in the main analysis does not include irredentist claims, that is demands of non-core groups to secede from the current state and be transferred to a neighboring ethnically kin state. These claims mostly co-occur with independence claims. As an additional robustness test, we replicate the main analysis including 7 additional irredentist claim onsets. Unsurprisingly, the estimates in Table A11 and the event study plots in Figure A11 closely match the main results without irredentism.

## Testing Conditional Hypotheses

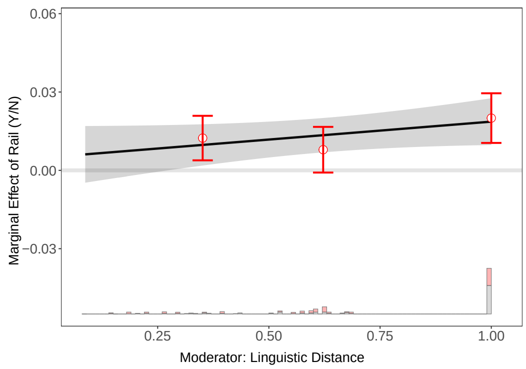
To test the conditional Hypotheses 2, we replicate the baseline model from Column 1 in Table 1 while interacting the railway access dummy with moderating variables coded at the segment- and country-year level. Figure 4 displays marginal effect plots along with the binning estimates as proposed by Hainmueller, Mummolo, and Xu (2019). Detailed results are presented in Appendix Tables A12 and A13.

Figure 4a tests whether the destabilizing effects of rails is stronger in ethnic segments that are culturally more distinct from the state-leading group (H2a). We calculate linguistic distance from the core group by matching the ethnic categories from our maps to the Ethnologue language tree. Interacting the rail treatment with linguistic distance yields a positive but merely weakly significant coefficient (Model 1 in Table A13). For example, separatist conflict took place both between linguistically similar groups such as Catalans and Spanish and distant ones such as Germans and Hungarians. 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.

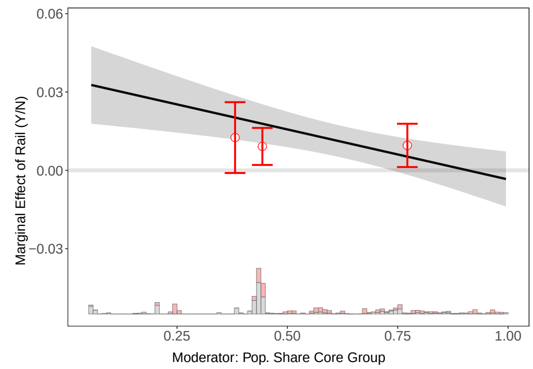
Figure 4b interacts the rail indicator with the country-year-level population share of the dominant national core group. Consistent with Hypothesis 2b, the interaction coefficient is negative and significant suggesting local railways are particularly likely to spur nationalist independence campaigns in countries with relatively small ruling groups. However, the binning plot in Figure 4b suggests that the significant linear interaction term is likely due to a small number of cases with particularly small core groups.[11] The binning coefficients show that there are no significantly different effects in the lowest, intermediate, and highest tertiles of the distribution of national core group's population share.

11. Note that our population measures underestimate the population size of Russians in the Russian Empire and Turks in the Ottoman Empire because of our geographical definition of Europe, which crops part of each group's population.

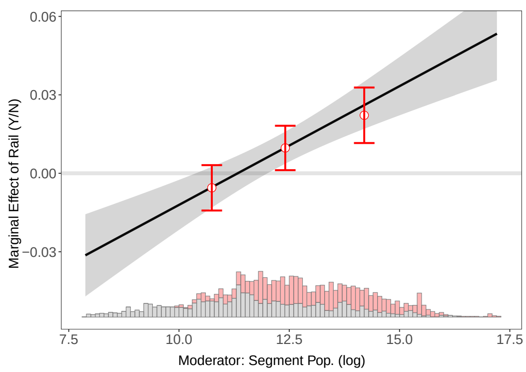
Figure 4: Marginal Effect Plots & Binning Estimates



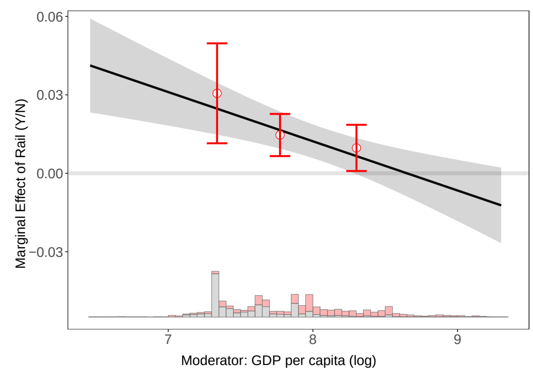

(a) Linguistic distance



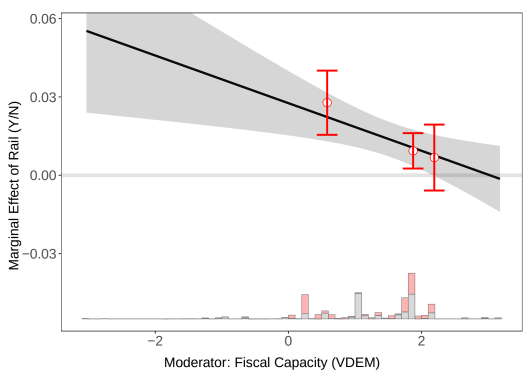

(b) Share core group



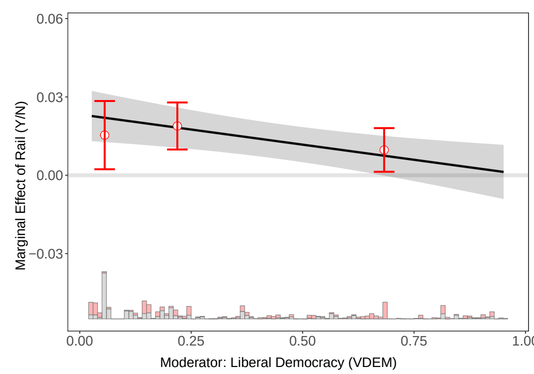

(c) Segment population




(d) Per capita GDP




(e) Fiscal capacity




(f) Liberal democracy

Note: The linear interaction estimates derive from models in Table A12, binned estimates from Table A13.

Figure 4c tests our argument about the non-core groups' opportunities to engage in separatism. The results reveal that railways mainly spur separatism in demographically large ethnic segments, in line with hypotheses H2c. Examples of large ethnic segments that mobilize are Belorussians, Poles and Ukrainians in Russia, and Czechs, Hungarians and Italians in Austria-Hungary. In contrast, railroad access has a *negative* effect in very small ethnic segments, in which it is likely more difficult to stage a separatist movement against the forces of state and market integration.

To test H2d and H2e, we rely on per capita GDP and fiscal capacity measures from the historical V-Dem data (Coppedge et al. 2016). The negative and significant linear interaction with per capita income in Figure 4d suggests that our findings are driven by relatively poor and arguably less industrialized country-years in the sample, thus confirming H2d. Similarly, the binning estimates for fiscal capacity in Figure 4e 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 H2e. The cases of separatism in less developed states with lower fiscal capacity mostly fall in the Russian and Ottoman Empires and in their successor states.

Finally, the interaction term with the V-Dem liberal democracy score (Coppedge et al. 2016) is negative and significant (Figure 4f). However, the binning estimates reveal that, if anything, the effect is highest at low-to-intermediate values of liberal democracy, which mostly occur in the Ottoman and Russian Empires during the second half of the nineteenth century. While the rail effect in the most democratic tertile is significantly smaller in the intermediate one, it is not significantly lower than among observations in the lowest tertile.

## Exploring Causal Mechanisms

Finally, we attempt to separate the three mechanisms through which railway construction affect center-periphery bargaining and separatist mobilization as outlined in the theory. Thus, we compute railway-based proxies for (1) segments' economic market access (H3a) as their average travel time towards large cities (logged due to its skew),[12] (2) local state's reach (H3b) as

12. See Appendix A10, Table A14 for equivalent results after log-transforming all moderating variables.

Table 3: 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. All models control for the number of past conflicts and peace years indicators. Segment clustered standard errors in parentheses. *+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

the inverted average travel time to the capital, and (3) their internal connectivity (H3c) as the inverted[13] average travel time among their inhabitants.[14] In the main analysis, we use time-invariant population data from before the arrival of railroads to avoid biases from endogenous population developments. However, our results remain consistent when we compute all network statistics using time-variant population data (Tables A15 and A16 in Appendix A10). Table 3 shows TWFE models of separatism where these variables replace our baseline railway dummy variable. Given the continuous nature of our network measures, we cannot estimate difference-in-difference models as in the main analysis, thus requiring stronger assumptions on the absence of (time-variant) omitted variables and reverse causality.

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 come with substantive reductions in the likelihood of separatist mobilization as predicted by Hypothesis

13. Inversions are computed as $x_{inv} = min(x) + max(x) - x$ to ensure that larger values capture greater state reach and internal connectivity.

14. Lacking precise data, travel times are computed assuming constant speeds of 60km/h on railroads and 6km/h elsewhere. See Appendix A4.

H3b. Improving state reach by one standard deviation leads to a decrease in the risk of separatism onsets by .79 percentage points or 70 percent of the average risk. 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 H3c. Increasing segments' internal connectivity by one standard deviation comes with an increase in the risk of separatism onsets by .34 percentage points.[15] The negative and borderline significant coefficient of *National Market Access* turns substantively small and statistically 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. Additional analyses in Appendix A10 show that these results are robust to adding country-year fixed effects (Table A18) and controlling for leads of the independent variable that capture potential reverse causality (Tables A19 and A20).

These results provide stronger support for the political and mobilization-related mechanisms M2 and M3 than for nation building via market integration and social communication (M1). Another interpretation 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.[16] The fact that our baseline analysis shows positive effects of the first railway link in a segment 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. Such resistance needs to occur before it is too late, i.e. after railway access and internal connectivity improve local mobilization capacity, but before the state assimilates peripheral populations (Deutsch 1953). In addition, a more selective indicator for

15. See Table A17 in Appendix A10 for results with standardized network measures that facilitate coefficient comparison.

16. Data limitations prevent us from exploring this possibility in more detail. For a study of heterogeneous effects of railroads on local population dynamics in Britain and Wales, see Bogart et al. (2022).

culturally distinctive direct rule of "nationalizing" states (Brubaker 1996) could yield different results.

## Conclusion

Modern transportation infrastructure is conventionally seen as having strengthened European state- and nation-building. Expanding railway networks boosted centralizing states' infrastructural power and enabled increasingly direct forms of governance, while spurring economic change, urbanization, and social contact over increasing distance.

Extrapolating from Weber's (1976) study of nationalizing France, many social scientists expect these changes to have strengthened national cohesion well beyond the French case. Yet, this paper shows that, if anything, railway construction in ethnic minority regions tended to threaten the integrity of European states and empires. Our analyses suggest that separatism became more likely after territories inhabited by non-leading ethnic groups were connected to the state's railroad network. Our conditional analysis reveals some structural dimensions that hindered national integration in multi-ethnic states, especially in Eastern Europe. Large minority groups, small population shares of state-leading groups, weak levels of state capacity and per capita income posed formidable challenges for state centralization and top-down nation-building. Thus, the French experience appears more as an exception than a paradigmatic case of nation building in Europe.

We also show how the aggregate effects of railroad access mask varying effects of the networks' overall structure. Results from our analyses of causal mechanisms suggest that separatism becomes more likely where railroads facilitate mobilization by improving internal connectivity of peripheral ethnic regions but less likely where it brings such regions closer to the state's capital. National market access, however, does not seem to make a difference.

Railway construction was only one, though arguably the most important, vector of modernization in Europe from the 19th century through the mid-20th century. 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. There is a growing research agenda analyzing 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 is needed before applying our findings to settings beyond the classical cases of European nation building.

## On Ethics and Conflict of Interest

The authors declare no ethical issues or conflicts of interest in this research. This reseach was funded by the Advanced ERC Grant 787478 NASTAC, "Nationalist State Transformation and Conflict."

## 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. 2023. "The 'weight' of territorial issues: Evidence from Catalonia, Scotland, and Northern Ireland." *Comparative Politics* 56 (1).

Barbour, Stephen, and Cathie Carmichael. 2000. *Language and Nationalism in Europe.* Oxford: Oxford University Press.

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.

Bogart, Dan, Xuesheng You, Eduard J Alvarez-Palau, Max Satchell, and Leigh Shaw-Taylor. 2022. "Railways, divergence, and structural change in 19th century England and Wales." *Journal of Urban Economics* 128:103390.

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.

Chaisemartin, Clément de, and Xavier D'Haultfœuille. 2022. "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey." *The Econometrics Journal* 26, no. 3 (June): C1–C30.

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.

Coppedge, Michael, John Gerring, Staffan I Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, M Steven Fish, Adam Glynn, Allen Hicken, et al. 2016. "V-Dem Dataset."

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

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.

Fogel, Robert William. 1964. *Railroads and American Economic Growth.* Baltimore: Johns Hopkins University Press. Book.

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.

Garcia-Alvarez, Jacobo. 1998. "Substate nation-building and geographical representations of the "other" in Galicia, Spain (1860–1936)." *Finisterra* 33 (65).

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. 2021. "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.

Goina, Calin. 2005. "How the State Shaped the Nation: an Essay on the Making of the Romanian Nation." *Politics and Society* 8.

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.

Hutchinson, John. 1987. "Cultural Nationalism, Elite Mobility and Nation-Building: Communitarian Politics in Modern Ireland." *The British Journal of Sociology* 38 (4): 482–501.

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. 2024. "A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data." *American Journal of Political Science* 68 (1): 160–176.

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

Malesevic, Sinisa. 2012. "Did Wars Make Nation-States in the Balkans?: Nationalisms, Wars and States in the 19th and early 20th Century South East Europe." *Journal of Historical Sociology* 25 (3): 299–329.

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.

Martí-Henneberg, Jordi. 2021. "From State-Building to European Integration: The Role of the Railway Network in the Territorial Integration of Europe, 1850–2020." *Social Science History* 45 (2): 221–231.

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

Minahan, James. 2001. *One Europe, Many Nations: A Historical Dictionary of European National Groups.* Westport, Connecticut: Greenwood.

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.

Müller-Crepon, Carl, Guy Schvitz, and Lars-Erik Cederman. 2023. "Shaping states into nations: The effects of ethnic geography on state borders." *American Journal of Political Science.*

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.

Paglayan, Agustina S. 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.

Pruett, Lindsey. 2023. "Resisting the Blood Tax: Coercive Capacity, Railroads and Draft Evasion in Colonial West Africa." *Journal of Politics, forthcoming.*

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.

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe. 2023. "What's trending in difference-in-differences? A synthesis of the recent econometrics literature." *Journal of Econometrics.*

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.

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

## *Supplementary Materials (Online)*

**Yannick I. Pengl** *ETH Zurich*
yannick.pengl@icr.gess.ethz.ch

**Carl Müller-Crepon** *LSE*
c.a.muller-crepon@lse.ac.uk

**Roberto Valli** *Princeton University*
roberto.valli@princeton.edu

**Lars-Erik Cederman** *ETH Zurich*
lcederman@ethz.ch

**Luc Girardin** *ETH Zurich*
luc.girardin@icr.gess.ethz.ch

September 9, 2024

## Contents



# 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.[1]

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

1. 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.[2]

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.[3]

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.

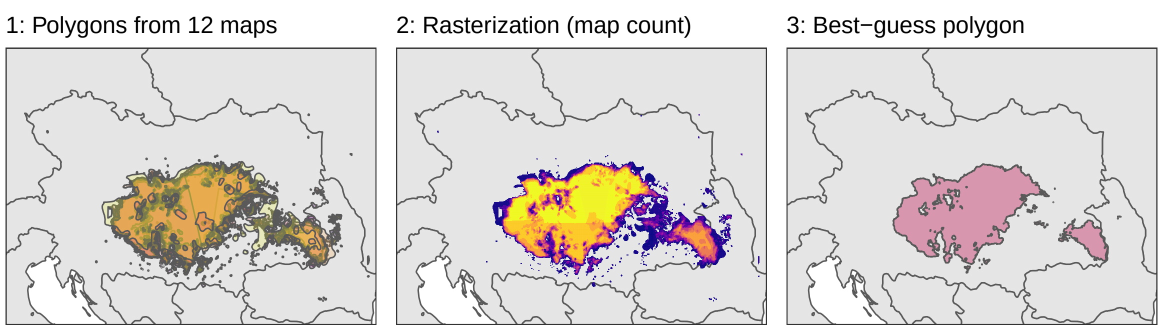



Figure A1: Constructing ethnic best-guess polygons: Hungarian example

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. While there is no perfect solution to this issue, we run robustness checks using temporally stable grid cells as units of analysis.

2. 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.

3. 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.

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.

## 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, 1876, 1881, 1884, 1885, 1991, 1995, 1901 from the Rumsay historical map collection.[4] 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 kilometer 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 5 and 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. The plot subtitles report that more than 80 percent of points are nested within 5 kilometers from the Austro-Hungarian lines, and more than 90 percent lie within 10 kilometers.

Figure A3 also provides the trends of these statistics over time. Plot A3a indicates that the distance of the rail lines contained in the two railway datasets is at its highest in 1864 with about 6 kilometers on average, and it decreases over time. As a result, the share of points along RShapes within 5 and 10 kilometers from the Austro-Hungarian lines increases over time.

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.

4. See https://www.davidrumsey.com/.

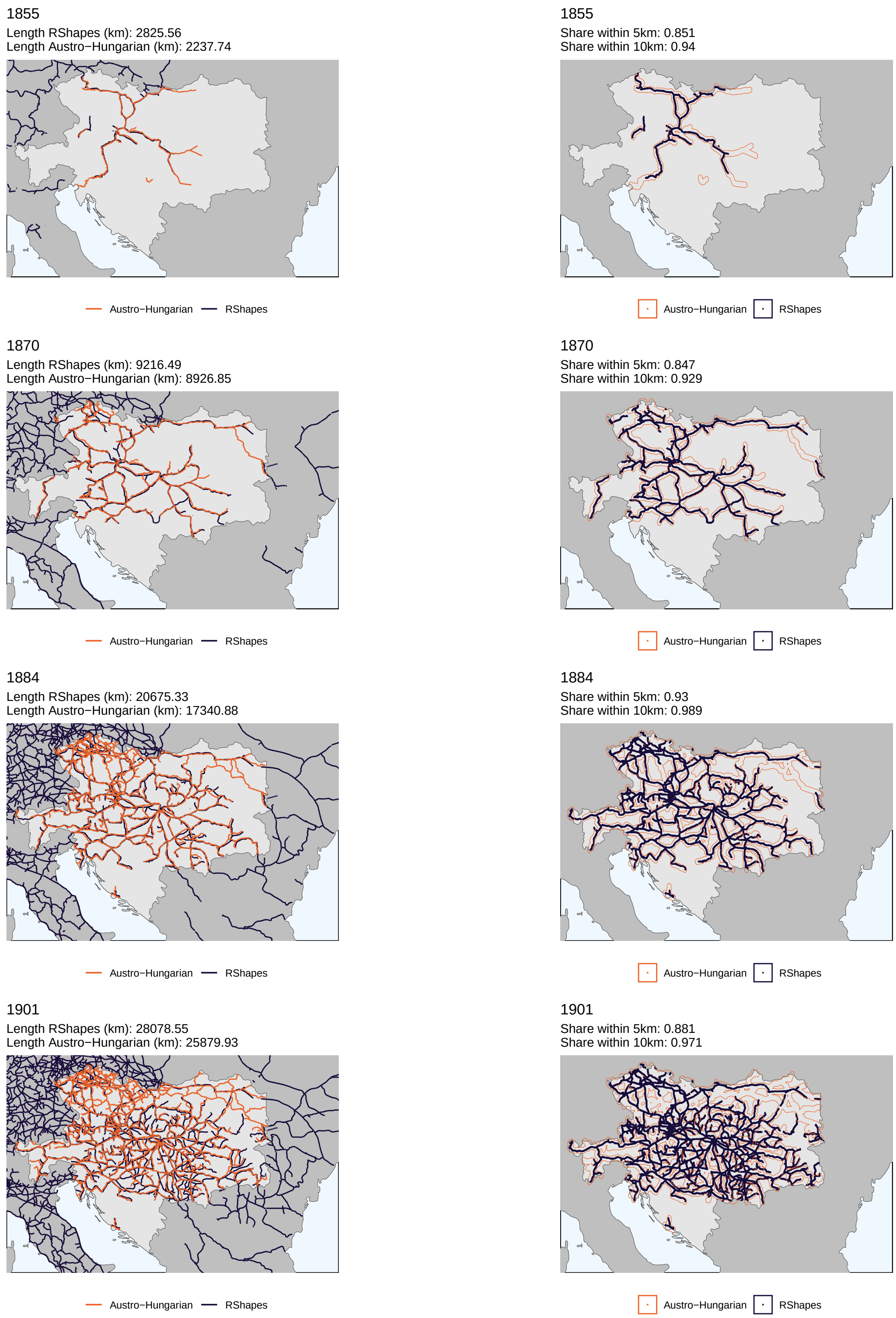
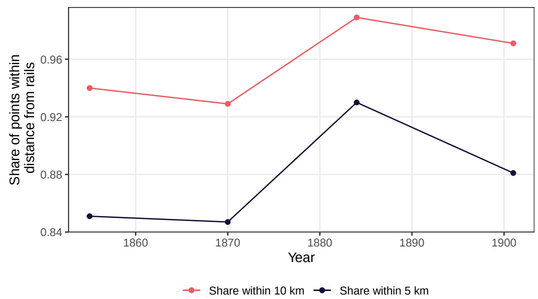
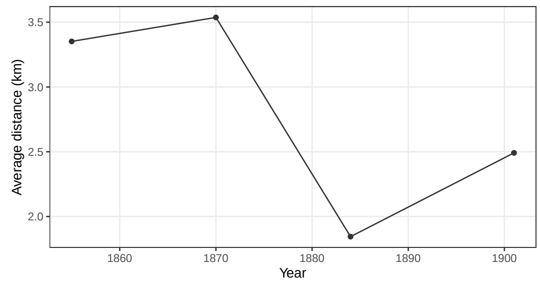



Figure A2: Comparison of RShapes and Austro-Hungarian railway data.




(a) Average distance between RShapes and Austro-Hungarian lines.




(b) Share of RShapes points within 5 or 10km from Austro-Hungarian lines.

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

# A3 Descriptive statistics

Table A1: Descriptive statistics

| | Min | Mean | Median | Max | Std. dev. |
|---|---|---|---|---|---|
| Combined outcome* | 0.000 | 1.115 | 0.000 | 100.000 | 10.500 |
| Successful secession* | 0.000 | 0.131 | 0.000 | 100.000 | 3.613 |
| First claim* | 0.000 | 0.569 | 0.000 | 100.000 | 7.522 |
| Civil war* | 0.000 | 0.438 | 0.000 | 100.000 | 6.606 |
| Rails (Y/N) | 0.000 | 0.512 | 1.000 | 1.000 | 0.500 |
| First railway year | 1835.000 | 1870.176 | 1868.000 | 1921.000 | 19.739 |
| National Market Access | -16.498 | -4.545 | -3.096 | 4.205 | 4.656 |
| State Reach | 0.000 | 344.898 | 381.408 | 421.340 | 93.884 |
| Internal Connectivity | 0.000 | 188.765 | 194.527 | 205.476 | 19.981 |
| Ling. Dist to Core | 0.087 | 0.736 | 0.684 | 1.000 | 0.281 |
| Pop. Share Core Group | 0.056 | 0.525 | 0.443 | 0.995 | 0.229 |
| Group Population (log) | 7.876 | 12.432 | 12.409 | 17.209 | 1.743 |
| GDP per capita (log) | 6.460 | 7.816 | 7.772 | 9.302 | 0.469 |
| Fiscal Capacity (VDEM) | -3.034 | 1.252 | 1.504 | 3.178 | 0.828 |
| Liberal Democracy (VDEM) | 0.027 | 0.336 | 0.218 | 0.951 | 0.282 |

* Note: The outcome is multiplied by 100 to improve legibility.

## A3.1 Descriptives of outcome variables

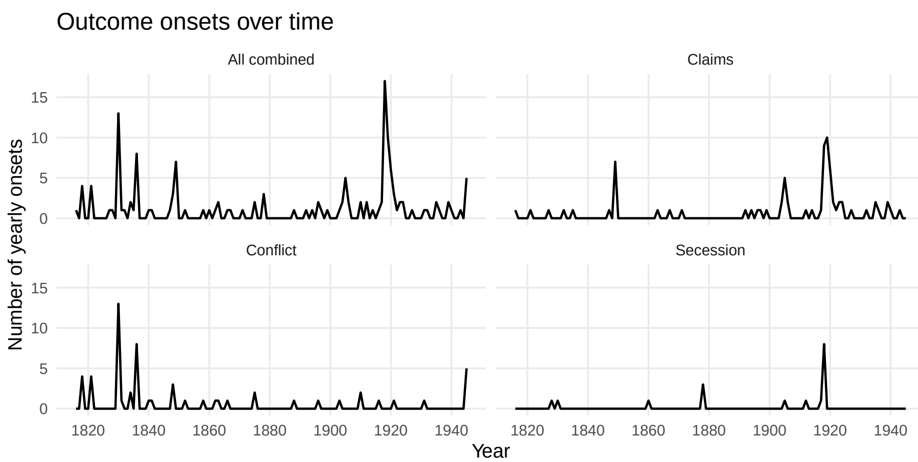
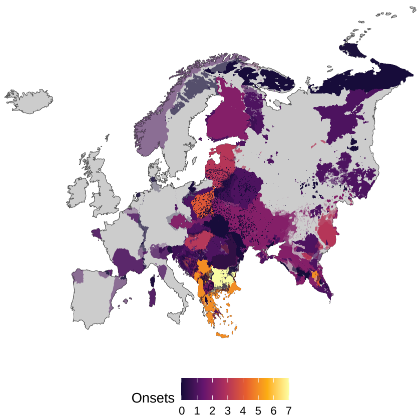
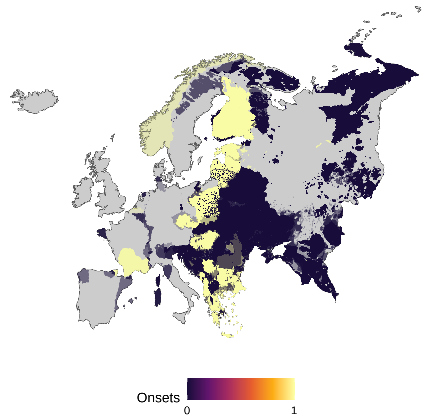


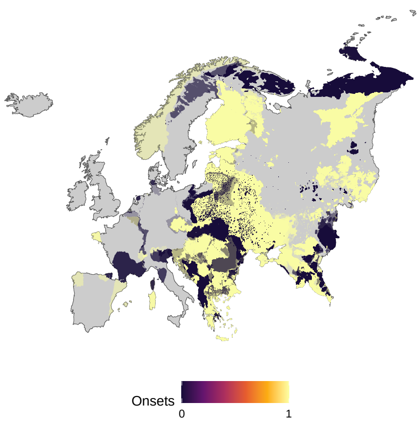

Figure A4: Temporal trends for outcome variables.



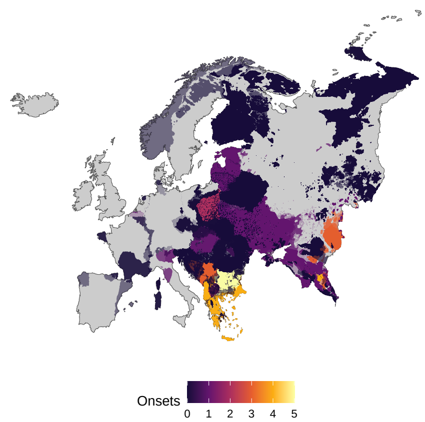

(a) All combined




(b) Claims




(c) Conflict




(d) Secession

Figure A5: Spatial maps of outcome variables.

# A4 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,[5] and 60 km/h for rail travel. The latter is close to the maximum average long-distance speeds

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

achieved by steam-powered trains in 19$^{th}$ century France, Italy, and the United Kingdom. While not entirely accurate, we currently lack more detailed data on changes in speeds over time and, even more challenging, variation in speeds by railroad line.[6]

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. It is then inverted,[7] to ensure that high values point to high levels of state capacity (i.e., low travel times).

The **national 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.[8] Travel times to different cities are weighted by market size (i.e., city population, from Buringh 2021) and distant cities are weighted down by a trade elasticity parameter based on travel times using parameters estimated by Donaldson and Hornbeck (2016). In particular, we compute the market access of a grid cell $i$ as

$$MA_i = \sum_{c=1}^{C} w_c \, \delta(i, c)^{-3.8}$$

, where c indexes cities in the same state as $i$ with a population size (weight) $w_c$ located at a distance of travel time $\delta(i, c)$ from grid cell $i$.[9]

We again aggregate cell-level market access values to ethnic segment-years by taking the population-weighted 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 rails built elsewhere that increase the overall connectivity within national networks.

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. It is then inverted,[10] to ensure that high values point to high levels of state capacity (i.e., low travel times).

The use of time-invariant population data for the main analysis limits the precision of our measures, it has the strong advantage that demographic developments caused by factors *other* than railroads do not affect our analysis. For example, all three measures of national market access, state reach, and internal connectedness can change due to changes in local demography, changes which are likely driven by a host of factors that are not related to railway networks and the economic modernization they bring about, but may cause conflict, thereby biasing our results. We do, however, conduct a set of analysis using time-variant population data from Goldewijk, Beusen, and Janssen (2010, for grid cells) and Buringh (2021, for cities) to construct all three measures. The results robust to this change and are presented and discussed in Appendix A10.

6. Introducing temporal variation in speeds (within reasonable limits) would not, generally, affect comparisons in our analytical framework much as most of their effect would be soaked up by our year and country-year fixed effects. Introducing measures of track quality would likely improve the precision of our measures and avoid attenuation bias.

7. Using the following formula: $x_{inv} = min(x) + max(x) - x$

8. The measure is closely related to Schürmann and Talaat's (2002) measure of peripherality.

9. We add 1 hour to all travel times to avoid division by 0.

10. Using the following formula: $x_{inv} = min(x) + max(x) - x$

## A5 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 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.

## A6 Railroad simulation

**Intuition:** Our simulation procedure starts from the intuition that, in the absence of local policing or external military goals or economic motivations, 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 A6, 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 resulting 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. We choose to temporally fix the population distribution at its estimate for 1830 to preclude that changes in the population distribution – which might be caused by observed railroads or other proximate causes of conflict – affect and potentially bias our simulation. 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. Our use of stringent ethnic segment and (state-)year fixed effects control for each of these factors.

The simulation is different from others' in the literature (Bogart et al. 2022; Faber 2014; Jedwab, Kerby, and Moradi 2017) in that it does not presuppose any fixed set of *nodes* that must be connected to the railway network but instead lets the algorithm find appropriate nodes to connect. While the latter approach works well for identifying the local effects of localities' access to the railway, it leaves the overall structure of the network fixed and is therefore not suitable for our approach. In addition, we are interested in the spatial evolution of the network across many years which allows us to instrument changes in railway access within ethnic segments *over time*. In contrast, the above mentioned studies focus on identifying networks' structure at a given point in time, which does not allow for capturing dynamic evolution across more than two periods.

**Technical details:** 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$^{\text{th}}$ 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.

# A7 Choice of estimators

This paper analyzes a setting in which the construction of railways can be analyzed as a non-reversible treatment with staggered adoption and a control group mostly composed of not-yet-treated units. The econometric literature in political science and economics identified several challenges to traditional estimation techniques in this type of settings and proposed several estimators that address these challenges. We address this literature and its empirical implications by estimating comparable models from estimators that make different assumptions and estimation choices in order to assess the robustness of our findings.

Our baseline model is the traditional two-way fixed effects (TWFE) estimator, originally proposed to estimate average treatment effects on the treated in settings with contemporary treatment adoption (Angrist and Pischke 2009), and recently criticized for producing biased estimates in settings with staggered adoption (e.g. Goodman-Bacon 2021; Sun and Abraham 2021; Chaisemartin and D'Haultfœuille 2020; Callaway and Sant'Anna 2021). All alternative solutions to the TWFE estimator address i) the problematic comparisons implicit to TWFE estimates, which sometimes use already-treated units as control observations for later cohorts, and ii) intransparent and sometimes counterintuitive weights given to cohort-specific treatment effect estimates in TWFE, whereby some units' estimated effect is given negative weight. (For an overview of the literature, see Roth et al. (2023) and Chaisemartin and D'Haultfœuille (2022).)

We choose two alternatives to TWFE that we believe represent the best choice for our empirical setup. First, we use the two-stage DiD estimator proposed by Gardner (2021) and implemented in the `did2s` R package (Butts and Gardner 2021). The intuition of the method consists in using the residualized control units (after partialling out unit and time fixed effects, and the necessary control variables) to impute the counterfactual outcomes for the treated units in a first stage. The second stage regresses the observed and the imputed outcome variables on the treatment indicator with a simple linear regression. A major benefit of the `did2s` package is that it allows to interact the main treatment variable with other factors to study heterogeneous effects. Therefore, estimates based on the two-stage DiD are used in all the main results in the main paper, as well as the main robustness test.

Second, we employ the `fect` package proposed by Liu, Wang, and Xu (2024). It follows a similar imputation approach to Gardner (2021), yet with significant differences. To begin with,

Table A2: Alternative panel data estimator: `fect` package

| | 100 × Secession, Terr. CW or Claim | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Rails (Y/N) | 1.622** | 1.625*** | 1.625** | 1.385* |
| | (0.512) | (0.483) | (0.545) | (0.688) |
| Civil war history | −6.789*** | −6.789*** | −6.789*** | −6.945*** |
| | (1.740) | (1.652) | (1.563) | (1.479) |
| Time since civil war | −0.095 | −0.095 | −0.095 | −0.104 |
| | (0.093) | (0.097) | (0.089) | (0.089) |
| Method | FE | IFE | MC | CFE |
| Segment FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Time since ind. or aut. claim | 0.027* | 0.027** | 0.027* | 0.027 |
| | (0.014) | (0.009) | (0.012) | (0.023) |
| Country-Year FE | No | No | No | Yes |

*Notes:* The unit of analysis is the ethnic segment year. State-leading segments and segments smaller than 2000 sqkm dropped. Method acronyms: 'FE' = two-way fixed effects; 'IFE' = interactive fixed effects; 'MC' = matrix completion; 'CFE' = complex fixed effects. Segment clustered standard errors in parentheses. *+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

it allows for different ways to impute counterfactual outcomes with interactive fixed effects and matrix completion models, which borrow from the fields of computer science and factor analysis. Additionally, `fect` computes the period-wise unit treatment effects as simple differences in means between observed and imputed outcomes, making fewer assumptions about effect linearity. Finally, the `fect` approach is more robust to temporal effect spillovers that might bias TWFE estimates (Liu, Wang, and Xu 2024).

Table A2 reports average treatment effect estimates from four models similar to the main results in Table 1 (article body). Model 1 uses a specification in which the counterfactual outcomes of treated units are predicted based on the trends of control units net of fixed effects for segments and years. Models 2 and 3 respectively use an interactive fixed effects method similar to generalized synthetic controls (Xu 2017), and a matrix completion method (see Athey et al. 2021) to predict counterfactual outcomes. Both methods are more flexible than TWFE in capturing heterogeneous temporal trends in the control group, and therefore might produce better counterfactuals. Model 4 uses a complex fixed effects estimator that allows to add country-year fixed effects on top of segment and year fixed effects, thereby resembling more Columns 2 and 4 in Table 1 (article body). Across all models, we obtain consistently similar positive and statistically significant estimates in line with the main results. Moreover, in line with Table 1, estimates with country-year fixed effects in Model 4 are smaller in magnitude than the ones with segment and year fixed effects.

Finally, we note that our data structure, does not allow the use of alternative estimators that require the presence of never treated units (of which there are exceedingly few in our setting) such as the ones introduced by Sun and Abraham (2021), Goodman-Bacon (2021), and Wooldridge (2022) and implemented by McDermott (2023) and Butts and Gardner (2021), nor those limited to two time periods (Sant'Anna and Zhao 2020).

# A8 Robustness tests

## Results with country-year fixed effects

### *Event study*

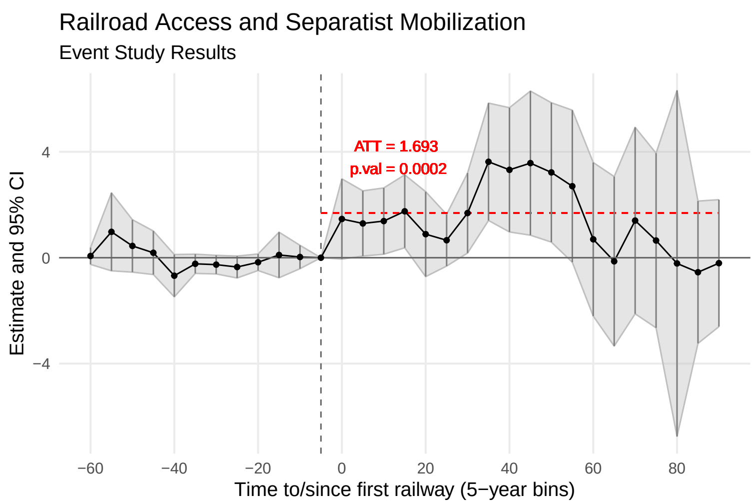



Figure A6: Event study plot
(ATT estimates based on Column 4 of Table 1 in the article)

### *Instrumental variable analysis*

Table A3: 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. All models control for the number of past conflicts and peace years indicators. Segment clustered standard errors in parentheses. *+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

## Sample definition

### *Limiting sample to treatment variation*

The main independent variable *Rail (Y/N)* is based on a dynamic railway network between 1834 and 1922. In our main models, we allow outcomes to unfold past 1922 to capture longer-time effects of railways construction that we would miss by censoring the outcome with the treatment variation. However, as an additional robustness test this section provides a complete set of results in which the data stops in 1922, including the DiD and event-study models, mechanism, and heterogeneity analysis.

Table A4: 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 A5: 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) | (1.466) | (1.404) |
| 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 A6: Railroads and Separatism (1816-1922)

| | 100 × Secession, Terr. CW or Claim | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Rails (Y/N) | 1.472*** | 1.057** | 2.610*** | 2.179*** |
| | (0.366) | (0.352) | (0.593) | (0.489) |
| 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 | 1.214 | 1.214 | 1.156 | 1.141 |
| Observations | 10 379 | 10 379 | 9951 | 8415 |

*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: Network Structure and Causal Mechanisms

| | 100 × Secession, Terr. CW or Claim | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| National Market Access | −0.141 | | | 0.017 |
| | (0.101) | | | (0.088) |
| State Reach | | −0.008* | | −0.009** |
| | | (0.003) | | (0.003) |
| Internal Connectivity | | | 0.018* | 0.019** |
| | | | (0.007) | (0.007) |
| Segment FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Mean DV | 1.241 | 1.214 | 1.214 | 1.241 |
| Observations | 10 072 | 10 379 | 10 379 | 10 072 |

*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. *+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

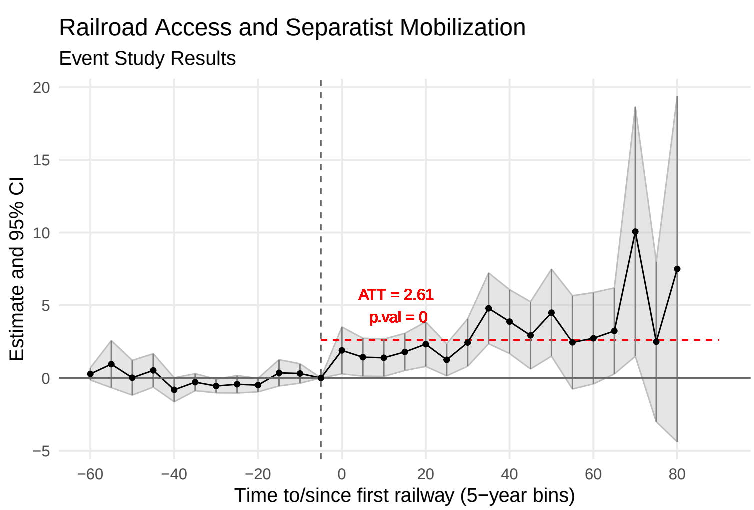



Figure A7: Event study plots (1816-1922)
(ATT estimates based on Column 3 in Table A6)

## Disaggregating the separatism outcomes

Table A8: 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. All models control for the number of past conflicts and peace years indicators. Segment clustered standard errors in parentheses. *+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

Table A9: 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. All models control for the number of past conflicts and peace years indicators. Segment clustered standard errors in parentheses. *+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

Table A10: 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. All models control for the number of past conflicts and peace years indicators. Segment clustered standard errors in parentheses. *+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

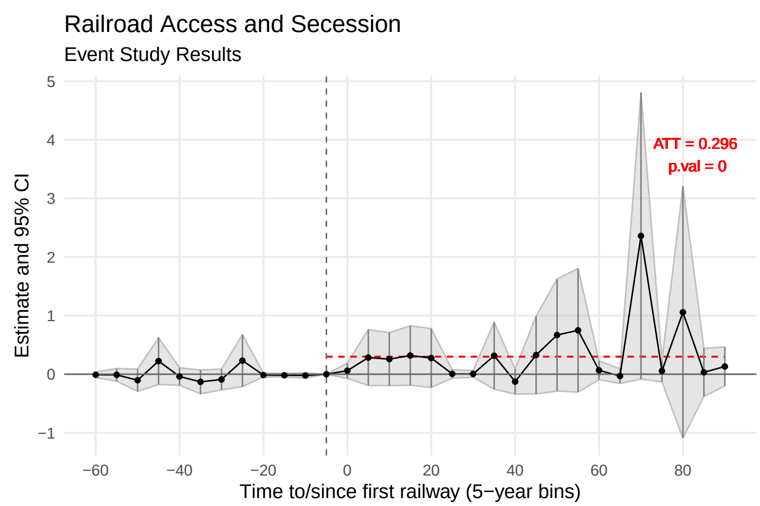



Figure A8: Event study plots: Secession
(ATT estimates based on Column 3 in Table A8)

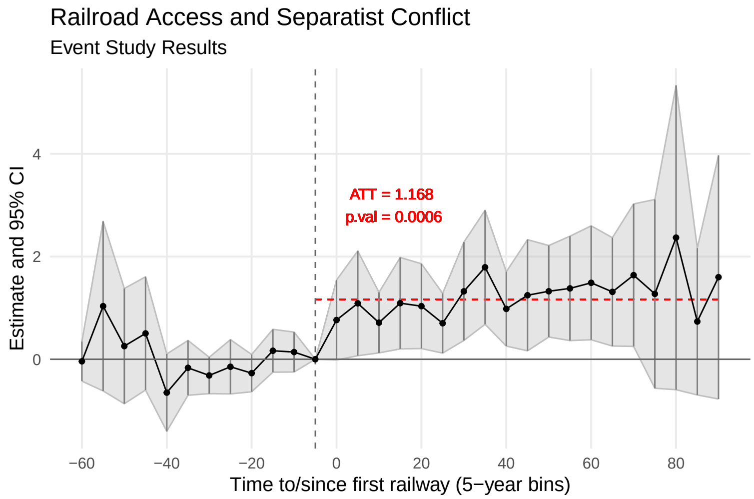



Figure A9: Event study plots: Separatist conflict
(ATT estimates based on Columns 3 in Table A9)

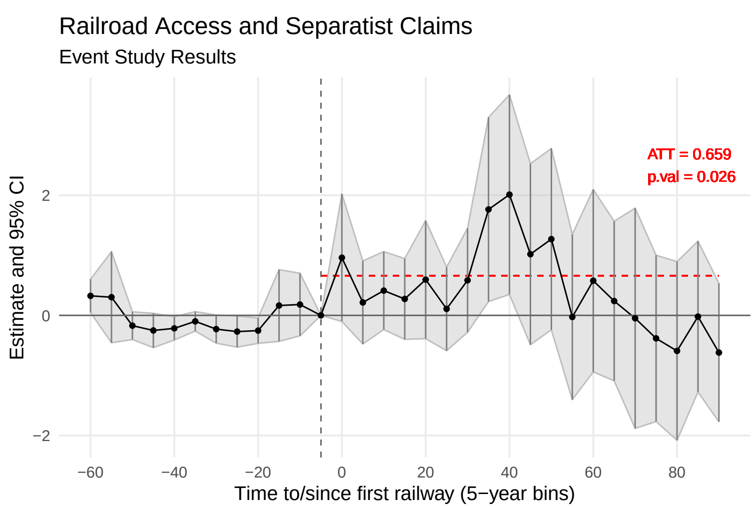



Figure A10: Event study plots: Autonomy and independence claims
(ATT estimates based on Column 3 in Table A10)

## A8.1 Including irredentism

Table A11: Railroads and Separatism or Irredentism (1816-1945)

| | 100 × Secession, Terr. CW or Claim (incl. Irredentism) | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Rails (Y/N) | 1.567*** | 1.157** | 1.751** | 1.596** |
| | (0.388) | (0.368) | (0.536) | (0.495) |
| 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 | 1.199 | 1.199 | 1.127 | 1.12 |
| 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. All models control for the number of past conflicts and peace years indicators. Segment clustered standard errors in parentheses. *+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

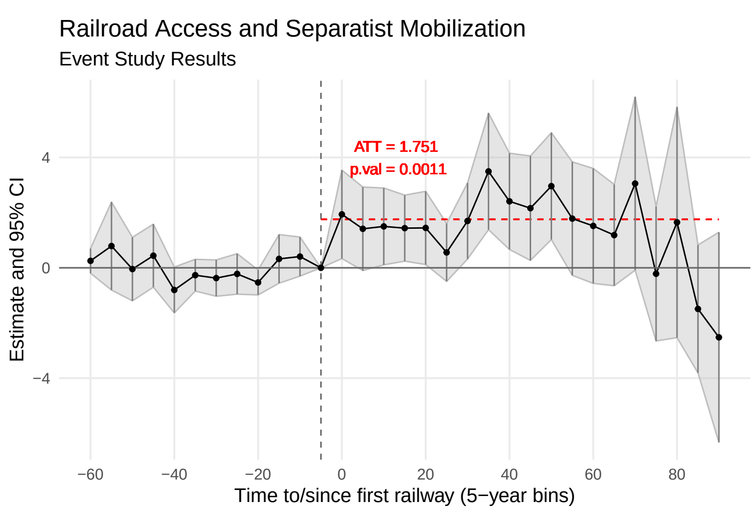



Figure A11: Event study plots: Including irredentism
(ATT estimates based on Column 3 in Table A11)

# A9 Conditional effects: regression tables

Table A12: Separatism: Binned Interaction Models

| | 100 × Secession, Terr. CW or Claim | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 1st tertile | 0.012** | 0.013+ | −0.006 | 0.031** | 0.028*** | 0.015* |
| | (0.004) | (0.007) | (0.004) | (0.010) | (0.006) | (0.007) |
| 2nd tertile | 0.008+ | 0.009* | 0.010* | 0.015*** | 0.009** | 0.019*** |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.003) | (0.005) |
| 3rd tertile | 0.020*** | 0.010* | 0.022*** | 0.010* | 0.007 | 0.010* |
| | (0.005) | (0.004) | (0.005) | (0.004) | (0.006) | (0.004) |
| Observations | 13 007 | 13 007 | 13 007 | 12 788 | 12 788 | 12 649 |
| Moderator | Ling. Dist to Core | Pop. Share Core Group | Group Pop. | GDP p.c. | Fiscal Cap. | Lib. Dem. |
| 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 |

*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 A13: Separatism: Linear 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. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

## A10 Additional mechanism analyses

The mechanism estimates based on measures of the structure of railroad networks remain consistent when we use time-variant population data to measure segments' average market access, state reach, and internal connectivity. Time-variant population data increases measurement precision, it risks bias from "baked-in" omitted variables that affect demographic developments. Results with year (Table A15) and country-year fixed effects (Table A16 show stable effects of state reach and internal connectivity. Counterintuitively, the effect of national market access turns positive and statistically significant when using only year fixed effects, a finding which is not robust to country-year fixed effects. This suggests its potential origin in bias introduced by the time-varying population data which the country-year fixed effects can partially account for.

Table A14: Network Structure: With and without Log-transform

| | 100 × Secession, Terr. CW or Claim | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| National Market Access (log) | −0.027 | 0.016 | | |
| | (0.076) | (0.102) | | |
| State Reach (log) | −1.137** | −1.193** | | |
| | (0.394) | (0.454) | | |
| Internal Connectivity (log) | 0.743* | 0.594+ | | |
| | (0.289) | (0.313) | | |
| National Market Access | | | 0.056 | 0.024 |
| | | | (0.046) | (0.041) |
| State Reach | | | −0.008** | −0.012*** |
| | | | (0.003) | (0.003) |
| Internal Connectivity | | | 0.016* | 0.013+ |
| | | | (0.007) | (0.007) |
| Segment FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | No | Yes | No |
| Country-year FE | No | Yes | No | Yes |
| Observations | 12 643 | 12 643 | 13 007 | 13 007 |
| Mean DV | 1.131 | 1.131 | 1.115 | 1.115 |

*Notes:* The unit of analysis is the ethnic segment year. State-leading segments and segments smaller than 2000 sqkm dropped. All models control for the number of past conflicts and peace years indicators. Segment clustered standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A15: Network Structure and Causal Mechanisms with Time-Variant Population Data

| | 100 × Secession, Terr. CW or Claim | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| National Market Access | 0.141* | | | 0.239** |
| | (0.069) | | | (0.071) |
| State Reach | | −0.009** | | −0.010*** |
| | | (0.003) | | (0.003) |
| Internal Connectivity | | | 0.015* | 0.022** |
| | | | (0.007) | (0.007) |
| Segment FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Mean DV | 0.963 | 1.115 | 1.115 | 0.963 |
| Observations | 11 732 | 13 007 | 13 007 | 11 732 |

*Notes:* The unit of analysis is the ethnic segment year. State-leading segments and segments smaller than 2000 sqkm dropped. All models control for the number of past conflicts and peace years indicators. Segment clustered standard errors in parentheses. *+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

Table A16: Network Structure and Causal Mechanisms with Time-Variant Population Data and Country-Year FEs

| | 100 × Secession, Terr. CW or Claim | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| National Market Access | −0.076 | | | 0.052 |
| | (0.101) | | | (0.103) |
| State Reach | | −0.013*** | | −0.012*** |
| | | (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 | 0.963 | 1.115 | 1.115 | 0.963 |
| Observations | 11 732 | 13 007 | 13 007 | 11 732 |

*Notes:* The unit of analysis is the ethnic segment year. State-leading segments and segments smaller than 2000 sqkm dropped. All models control for the number of past conflicts and peace years indicators. Segment clustered standard errors in parentheses. *+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

Table A17: Network Structure and Causal Mechanisms

| | 100 × Secession, Terr. CW or Claim | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| National Market Access | −0.664+ | | | −0.004 |
| | (0.387) | | | (0.351) |
| State Reach | | −0.742** | | −0.773** |
| | | (0.248) | | (0.253) |
| Internal Connectivity | | | 0.309* | 0.328* |
| | | | (0.140) | (0.132) |
| 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. All models control for the number of past conflicts and peace years indicators. Segment clustered standard errors in parentheses. All explanatory variables are standardized. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A18: Network Structure (Country-Year Fixed Effects)

| | 100 × Secession, Terr. CW or Claim | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| National Market Access | −0.091 | | | |
| | (0.104) | | | |
| State Reach | | −0.012*** | | −0.013*** |
| | | (0.003) | | (0.003) |
| Internal Connectivity | | | 0.014+ | 0.013+ |
| | | | (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. All models control for the number of past conflicts and peace years indicators. Segment clustered standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A19: Network Structure and Causal Mechanisms with 5-year Leads

| | 100 × Secession, Terr. CW or Claim | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| National Market Access | −0.072 | | | 0.092 |
| | (0.084) | | | (0.081) |
| State Reach | | −0.008** | | −0.009** |
| | | (0.003) | | (0.003) |
| Internal Connectivity | | | 0.017** | 0.020*** |
| | | | (0.006) | (0.006) |
| Δ National Market Access $t_{+5} - t_0$ | −0.037 | | | 0.109 |
| | (0.122) | | | (0.124) |
| Δ State Reach $t_{+5} - t_0$ | | −0.007 | | −0.010+ |
| | | (0.005) | | (0.006) |
| Δ Internal Connectivity $t_{+5} - t_0$ | | | 0.049 | 0.058 |
| | | | (0.039) | (0.040) |
| Segment FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Mean DV | 1.002 | 1.115 | 1.115 | 1.002 |
| Observations | 11 771 | 12 110 | 12 110 | 11 771 |

*Notes:* The unit of analysis is the ethnic segment year. State-leading segments and segments smaller than 2000 sqkm dropped. All models control for the number of past conflicts and peace years indicators. Segment clustered standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A20: Network Structure and Causal Mechanisms with 5-year Leads and Country-Year FEs

| | 100 × Secession, Terr. CW or Claim | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| National Market Access | −0.012 | | | 0.131 |
| | (0.095) | | | (0.101) |
| State Reach | | −0.011*** | | −0.013*** |
| | | (0.003) | | (0.003) |
| Internal Connectivity | | | 0.014* | 0.016* |
| | | | (0.006) | (0.006) |
| Δ National Market Access $t_{+5} - t_0$ | 0.118 | | | 0.209 |
| | (0.164) | | | (0.182) |
| Δ State Reach $t_{+5} - t_0$ | | 0.000 | | −0.002 |
| | | (0.007) | | (0.008) |
| Δ Internal Connectivity $t_{+5} - t_0$ | | | 0.036 | 0.039 |
| | | | (0.042) | (0.042) |
| Segment FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Mean DV | 1.002 | 1.115 | 1.115 | 1.002 |
| Observations | 11 771 | 12 110 | 12 110 | 11 771 |

*Notes:* The unit of analysis is the ethnic segment year. State-leading segments and segments smaller than 2000 sqkm dropped. All models control for the number of past conflicts and peace years indicators. Segment clustered standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

## References

Angrist, Joshua, and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* 1st ed. Princeton University Press.

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. 2021. "Matrix Completion Methods for Causal Panel Data Models." *Journal of the American Statistical Association* 116 (536): 1716–1730.

Bogart, Dan, Xuesheng You, Eduard J Alvarez-Palau, Max Satchell, and Leigh Shaw-Taylor. 2022. "Railways, divergence, and structural change in 19th century England and Wales." *Journal of Urban Economics* 128:103390.

Buringh, Eltjo. 2021. "The population of European cities from 700 to 2000: Social and economic history." *Research Data Journal for the Humanities and Social Sciences* 6 (1): 1–18.

Butts, Kyle, and John Gardner. 2021. *did2s: Two-Stage Difference-in-Differences Following Gardner (2021).*

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.

Chaisemartin, Clément de, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110 (9): 2964–2996.

———. 2022. "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey." *The Econometrics Journal* 26, no. 3 (June): C1–C30.

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.

Faber, Benjamin. 2014. "Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System." *Review of Economic Studies* 81 (3): 1046–1070.

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

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.

Jedwab, Remi, Edward Kerby, and Alexander Moradi. 2017. "History, path dependence and development: Evidence from colonial railways, settlers and cities in Kenya." *The Economic Journal* 127 (603): 1467–1494.

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

Liu, Licheng, Ye Wang, and Yiqing Xu. 2024. “A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data.” *American Journal of Political Science* 68 (1): 160–176.

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

McDermott, Grant. 2023. *etwfe: Extended Two-Way Fixed Effects.* R package version 0.3.5.

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.

Roth, Jonathan, Pedro H. C. Sant’Anna, Alyssa Bilinski, and John Poe. 2023. “What’s trending in difference-in-differences? A synthesis of the recent econometrics literature.” *Journal of Econometrics.*

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.

Sant’Anna, Pedro H. C., and Jun Zhao. 2020. “Doubly Robust Difference-in-Differences Estimators.” *Journal of Econometrics* 219:101–122.

Schürmann, Carsten, and Ahmed Talaat. 2002. “Towards a European peripherality index.” *Socio-Economic Applications of Geographic Information Science,* 255.

Sun, Liyang, and Sarah Abraham. 2021. “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects.” Themed Issue: Treatment Effect 1, *Journal of Econometrics* 225 (2): 175–199.

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.

Wooldridge, Jeffrey M. 2022. *Simple Approaches to Nonlinear Difference-in-Differences with Panel Data,* August.

Xu, Yiqing. 2017. “Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models.” *Political Analysis* 25 (1): 57–76.