# Is Secessionism Mostly About Income or Identity? A Global Analysis of 3,003 Subnational Regions*

Klaus Desmet
*SMU, NBER*, and *CEPR*

Ignacio Ortuño Ortín
*UC3M*

Ömer Özak
*SMU, IZA*, and *GLO*

August 7, 2023

**Abstract**

This paper analyzes whether the propensity to secede by subnational regions responds mostly to differences in income per capita or to distinct ethnolinguistic identities. We explore this question in a quantitative political economy model where people's willingness to finance a public good depends on their income and identity. Using high-resolution economic and linguistic data for the entire globe, we predict the propensity to secede of 3,003 subnational regions in 173 countries. We validate the model-based predictions with data on secessionist movements, state fragility, regional autonomy, and conflict, as well as with an application to the dissolution of the Soviet Union. Counterfactual analysis strongly suggests that identity trumps income in determining a region's propensity to secede. Removing ethnolinguistic identity differences reduces the average support for secession from 7.5% to 0.6% of the population.

*Keywords: Secessionism, Separatism, Identity, Political Economy, Language, Country Stability, Income*

*JEL Codes*: H77, P00, D70, D74, F02, F52, Z10

---

*Desmet: Department of Economics and Cox School of Business, Southern Methodist University, NBER and CEPR, kdesmet@smu.edu, ORCID 0000-0002-8206-845X. Ortuño Ortín: Department of Economics, Universidad Carlos III de Madrid, iortuno@eco.uc3m.es. Özak: Department of Economics, Southern Methodist University, IZA and GLO, ozak@smu.edu, ORCID 0000-0001-6421-2801. We thank Juan Díez Medrano for useful comments, and Angel Cuevas, Rubén Cuevas and José González Cabañas for help with the Wikipedia data. Ignacio Ortuño Ortín acknowledges financial support from the Spanish Ministry of Science, Innovation and Universities (Project PID2019-109157RB-IOO).

# 1 Introduction

Secessionist tensions pose a fundamental challenge for territorial governance across the globe, from Nagorno-Karabakh and Tibet in Asia, to Flanders and Catalonia in Europe, and Oromia and the Western Cape in Africa. In recent decades, referendums on independence have taken place in regions such as East Timor, Kurdistan, Puerto Rico, Quebec, Scotland, Somaliland, South Ossetia, and South Sudan, and pro-independence or regionalist political parties have held significant vote shares in regions such as the Basque Country and Lombardia. Although successful secessions have been uncommon since the breakup of the Soviet Union and Yugoslavia in the early 1990s, there are currently active secessionist movements in 54% of countries and 17% of subnational regions of the world. Some of these regions, such as Quebec, exhibit a distinct identity, eroding their sense of loyalty to the nation. Other regions, such as Lombardia, are relatively wealthy, spurring discontent among their population about subsidizing the rest of the country.

When evaluating the fundamental drivers of the demand for secession, a commonly held view is that identity differences matter, but only to the extent that income per capita differences are present (Sorens, 2005; Álvarez Pereira, Portos and Vourdas, 2018). That is, having a separate identity in terms of language, ethnicity, or religion is a necessary condition, but so is being relatively rich, with neither condition being sufficient (Gourevitch, 1979). This could explain why we see support for secession in the Basque Country and Flanders, relatively wealthy regions with a distinct language, but not in Galicia, a region with a different language but economically lagging, nor in Baden-Württemberg, a rich region but lacking a separate ethnolinguistic identity. However, it is not always obvious that both conditions need to be satisfied for secessionism to take hold. Regions such as Aceh and Tibet have significant pro-independence movements, in spite of their relatively low income per capita (Horowitz, 1981). These examples challenge the importance of high income per capita promoting secessionist sentiment.[1]

This paper asks how sensitive secessionism is to income per capita and ethnolinguistic identity differences. It tackles this question at an unprecedented global scale, for all first-level administrative regions of the world. Our analysis relies on a quantitative political economy model where people are less willing to contribute to a public good if they are richer or if they have an identity that is different from the rest of the country. Taking the model to the data for the entire world, we predict for 3,003 subnational regions the share of the population that favors secession. Counterfactual analysis suggests that identity

[1] In fact, the theory of internal colonialism in sociology and political science argues that lower income may lead to increased ethnic solidarity and facilitate ethnic political mobilization (Hechter, 1975; Díez Medrano, 1994).

trumps income in determining a region's propensity to secede. More specifically, removing linguistic identity differences between subnational regions lowers the support for secession from an average of 7.5% of the population to 0.6%, whereas changes in income per capita have almost no impact on the demand for secession. More than a necessary condition, having a separate ethnolinguistic identity is often a sufficient condition for separatism to emerge. This limits the scope of economic policies to stem territorial instability.

In our model, scale economies, income heterogeneity, and identity differences determine a country's stability. A country's geography is partitioned into subnational regions, each with a population size and an income level. A country's population is also partitioned into identity groups that may or may not coincide with the subnational regions. An identity group is a collection of individuals defined by a common trait such as a shared language or ethnicity.[2] Individuals have preferences over the consumption of private and public goods. The more an individual's identity differs from that of the rest of the population, the lower the utility she derives from public goods. Public goods are financed through a proportional income tax, decided by majority vote. As a result, richer individuals, as well as those with a distinct identity, end up paying a higher tax rate than they would prefer. Hence, subnational regions with higher income per capita and a different identity from the rest of the country experience on average greater support for independence. Pushing in the other direction are the scale economies associated with the provision of public goods, as well as within-region identity heterogeneity.

We quantify the model using high-resolution data on population, income per capita, and ethnolinguistic identity for the entire globe. For population, we use data from Landscan for the year 2000; for income per capita, the data come from G-Econ 4.0 and are also for the year 2000; for identity, we rely on a high-resolution database on language use developed by Desmet, Gomes and Ortuño Ortín (2020). Although the theory applies to any identity trait, we focus on language for three reasons. First, language has long been recognized as a major identity marker that differentiates populations (Fearon, 2003). Second, unlike other identity traits, we have detailed subnational language data for the entire globe. Third, other identity traits often strongly align with language: ethnicity largely coincides with language in most of Africa and Asia, and as we will show, secessionist regions with a distinct religion almost always also have a distinct language. For all countries of the world, we aggregate these high-resolution population, income, and linguistic data up to first-level administrative regions (e.g., U.S. states, German Länder, Russian oblasts, Colombian departments, Kenyan

[2] In an alternative use of terminology, we could refer to an individual's identity as the subnational region or the country she identifies with. If we were to use the term identity in that alternative sense, then an individual would endogenously choose her identity group. We return to this interpretation in more detail later.

provinces, Japanese prefectures). The model's parameters are then calibrated to match the territorial divisions of the current world map.

We use our calibrated model to estimate the share of each region's and each country's population in favor of secession. Within the top-10% group of regions with strongest support for independence, we find regions such as Tibet (China), Aceh (Indonesia), Lombardia (Italy), Okinawa (Japan), Tatarstan (Russia), and Catalonia (Spain). The most unstable countries for each of the six major continents are India, Italy, Papua New Guinea, South Sudan, Guatemala, and Bolivia. The share of the population that wants to secede is largest in South Asia (14%), followed by sub-Saharan Africa (13%), and it is smallest in North America, where less than 1% of the population favors secession.

We then use counterfactual analysis to investigate how sensitive secessionism is to income per capita and identity. First, we show that removing income per capita differences across subnational regions within countries has a limited effect on the demand for secession, whereas removing identity differences makes support for secession vanish. More specifically, starting from a baseline average regional support for secession of 7.5%, we see a drop of 6.9 percentage points in the absence of linguistic identity differences and no decline in the absence of income per capita differences. This suggests that identity is the main driver of the demand for separatism, whereas income per capita plays only a minor role. Second, we further assess how responsive secessionism is to changes in income per capita. Focusing on regions where at least 10% of the population supports independence, we find that income per capita would on average have to drop by 43% for its residents to give up on secessionism. That is, secessionism only weakens when separatist regions fall far behind economically. This points, again, to economic forces being of limited importance in shaping the drive for independence. Third, we analyze the effect of removing within-region linguistic identity differences, without changing the overall identity heterogeneity of the country. This leads to an almost tripling of the support for secessionism at the regional level: from 7.5% in the baseline to 20.0%. Hence, if subnational regions had a homogeneous identity, the model predicts that countries would become much more unstable. Once again, this shows that identity is an important force in shaping secessionism.

To further explore the robustness of our findings, we also consider an alternative calibration that targets public spending on defense and the regionalist vote share in a restricted set of developed countries. The counterfactual analysis under this alternative calibration confirms our main finding: identity is the main driver of secessionism. Removing income per capita differences has essentially no effect on the demand for secession, whereas removing linguistic identity differences lowers the support for secession from an average 6.4% of a region's population to 0.5% in this alternative calibration.

Taken together, these results strongly suggest that identity, more than income, is the key driver of secessionism. More than a necessary condition, having a distinct ethnolinguistic identity is often a sufficient condition to mobilize separatist sentiment. This limits the scope of economic policies in staving off secessionist threats. Instead, creating a common identity that cuts across existing ethnolinguistic cleavages is more likely to lead to territorial stability. Such policies were central to European nation-building efforts during the nineteenth century (Alesina, Giuliano and Reich, 2021). In Italy, for example, only a small proportion of its population spoke Italian at the time of unification in the mid-19th century. The adoption of a uniform and common language, based on the Florentine dialect, was viewed as an essential element in creating a unified Italian state. Similarly, at the time of the French Revolution, only 10% of the population spoke French fluently. The introduction of mass education under the *Loi Guizot* of 1833 has been credited with being the "ultimate acculturation process that made the French people French" (Weber, 1979).

To validate our calibrated model, we pursue three different strategies. First, we construct a global geographic database of active secessionist movements and analyze how well our model-generated measures of secessionism fit those data. More specifically, starting with a comprehensive list of 2,529 active secessionist movements taken from Wikipedia, we identify the first-level administrative regions associated with each movement. To get a measure of the importance of each secessionist movement, we use the number of Wikipedia page views in all languages from 2015 to 2020. We find that the predictions of the model align well with actual secessionist activity across the world, both across and within countries. Second, we compare our model predictions to data on the vulnerability of states to collapse, the autonomy of regional governments, and the intensity of conflict within countries. Reassuringly, in all cases there is a strong association with our measures of demand for secession. Third, using data from the end of the 1980s, we show that our calibrated model is able to account for the breakup of the Soviet Union. The model not only predicts that the country was unstable, but it also shows that the Soviet republics predicted to be most in favor of secession were the first ones to actually declare independence, and those least in favor of secession were the last ones to leave the union. In that sense, while we calibrate the model to the current world map, we are able to explain the perhaps most dramatic territorial change of the past half-century.

In our model, we associate an individual's identity with a given identity trait, such as language. If, instead, we were to associate an individual's identity with the territorial arrangement she prefers, then we could reinterpret our model as one with endogenous identity in the spirit of Shayo (2009, 2020).[3] More specifically, using the exact same model as before,

[3]This builds on the pioneering work of Akerlof and Kranton (2000) who emphasize the importance of

we would say that an individual chooses to identify with her subnational region rather than with her country if she would gain from her region seceding. As in Shayo (2020), an individual's choice of identifying with either her subnational region or her country would then depend on two factors: her material payoff in terms of public goods and her perceived proximity to the territory in terms of language. While we present our counterfactual exercises as exploring the relative importance of identity and income, under this alternative interpretation these same counterfactual exercises should be viewed as determining the relative importance of language and income. Ultimately, whether we refer to someone's language as her identity or not is a semantic choice.

Our paper is related to the theoretical, empirical, and quantitative literature on the stability of countries. Theoretical work on the size and number of nations has focused on the tradeoff between the benefits of scale economies and the costs of both preference heterogeneity (Alesina and Spolaore, 1997, 2003) and disagreements over fiscal redistribution (Bolton and Roland, 1997). This tradeoff is present in our framework as well: larger countries benefit from scale economies in the provision of public goods, but tend to suffer from greater regional differences in income per capita and identity.

Empirical work on country stability tends to conclude that both identity and income are necessary conditions for secessionism. In a series of studies on advanced economies, Sorens (2005, 2008) finds that economic concerns, more than ethnic identity, underlie the success of secessionist political parties. He argues that having a separate identity is only relevant if it can be mobilized to achieve economic goals.[4] Using a panel of 30 regions, Gehring and Schneider (2020) take a dynamic approach and show that an increase in regional income improves separatist party success. In an analysis of Western Europe, Álvarez Pereira, Portos and Vourdas (2018) emphasize a more balanced interaction between economic and cultural variables: support for autonomy and secession is greater in richer regions, but only to the extent that they are culturally differentiated. An exception is Horowitz (1981), who discusses cases of relatively poor regions that strive for independence because of their distinct identity. A shortcoming of most empirical studies is their limited geographical scope and their omission of intra-regional identity heterogeneity. Our paper overcomes this limitation by allowing for intra-regional heterogeneity, as well as covering the world.

There are a few quantitative papers on country stability, but they focus exclusively on Europe. Desmet et al. (2011) use a similar model to ours and quantify it to Western Europe, using genetic distances as a proxy for regional distinctiveness. Vanschoonbeek (2020) also

social identities for economic outcomes.

[4]See also Sambanis and Milanovic (2014) who study 48 decentralized economies and conclude that richer regions are more likely to demand greater autonomy.

analyzes how secession-prone different European regions are, but uses political distinctiveness in voting patterns as a proxy for regional differences. In contrast to these papers, we consider the entire globe, and hence include the many areas outside Europe where secessionist movements are strong. This global approach also allows us to make more general statements about the relative importance of income and identity as drivers of secessionism. Moreover, our analysis highlights the importance of within-region identity heterogeneity, a force that had remained unexplored by these previous papers. We find that the lack of a homogeneous identity in many subnational regions is a key impediment to secessionism.

Rather than focusing on scale economies coming from public goods, others have considered market access through trade as the main advantage of being part of a country. Comerford and Rodríguez Mora (2019) rely on a quantitative trade model to analyze the effect of Scotland, Catalonia and the Basque Country becoming independent nations, and estimate losses in real GDP ranging from 8.5 percent in Scotland to 16 percent in the Basque Country.[5] For the purpose of studying nation stability, we prefer to focus on a model with public goods, because market access through trade is in today's world no longer tightly related to country borders. While in the nineteenth century trade access was mostly achieved by removing country borders through empire building, in the post-WWII period gaining market access has mostly been achieved by signing trade agreements between countries (Gancia, Ponzetto and Ventura, 2022). Country borders are now less of an impediment to trade, as reflected by smaller countries being more open to international markets (Alesina, Spolaore and Wacziarg, 2000).

The rest of the paper is organized as follows. Section 2 proposes a simple model of a region's propensity to secede; Section 3 describes the data and the calibration; Section 4 reports the model-predicted propensity to secede for all first-level administrative regions of the world; Section 5 conducts counterfactual analysis to determine the relative importance of identity and income in driving the demand for secession; Section 6 assesses the model's performance; and Section 7 concludes.

# 2 A Simple Theory of the Propensity to Secede

In this section we introduce a political economy model that explores the relative importance of economic and identity differences as drivers of the demand for secession. Our simple framework is based on Desmet et al. (2011), and captures the trade-off between scale economies

[5] In related work, Gurevich et al. (2020) explore the role of language as an internal trade barrier. In particular, they estimate a gain in Canadian GDP of 1.22 percent if Quebec were to become fully bilingual in English.

and heterogeneity in determining the stability of a country. The scale economies come from the existence of public goods that are financed through taxes, and the heterogeneity comes from differences in income per capita and identity.

## 2.1 Setup

**Countries, subnational regions and identity groups.** A country $C$ is geographically partitioned into $R(C)$ subnational regions, indexed by $r$. There are $N(C)$ individuals, partitioned into $L$ identity groups, indexed by $\ell$ or $k$. Denote by $N(\ell, r, C)$ the number of individuals of group $\ell$ living in region $r$ of country $C$. Income per capita differs across regions but not within regions.[6] Denote by $y(r, C)$ the income per capita of an individual who resides in region $r$ of country $C$, and by $Y(C)$ the total income in country $C$.

We think of regions as administrative units, and of identity groups as collections of individuals defined by a common trait, such as language, ethnicity, or religion.[7] Often there is some overlap between regions and identity groups, but typically that overlap is not perfect. To give an example, take Canada and consider language as a group's identity marker. Most people in Quebec are French-speaking, and most people in New Brunswick are English-speaking, but both provinces have sizable minorities speaking the other language.

**Preferences over private and public goods.** An individual of group $\ell$ residing in region $r$ and country $C$ has quasi-linear preferences over private consumption $x(r, C)$ and public consumption $G(C)$ of the form

$$u(x(r, C), G(C), S(\ell, C)) = x(r, C) + \alpha S(\ell, C)^{\delta} G(C)^{\beta} \tag{1}$$

where $\alpha > 0$ and $\beta, \delta \in [0, 1]$, and $S(\ell, C)$ is the share of people in country $C$ that belong to identity group $\ell$. The idea is that an individual derives more utility from the public good, the greater the relative size of her identity group. To use the same example as before, to an English-speaking individual in Canada, public consumption is more valuable the greater

[6] Allowing for income differences within regions is theoretically a straightforward extension. However, because of data limitations on within-region income inequality, we do not explore this possibility here. Although luminosity data could be used in regions where the different identity groups live in geographically distinct areas, those data would not help us to measure income differences in regions where identity groups share the same geographic area. For work on income inequality between ethnic groups, see Alesina et al. (2016) and Huber and Mayoral (2019).

[7] We take an individual's identity group as given. Rather than associating an individual's identity group with her language or ethnicity, we could associate her identity group with either her subnational region or her country. Under this alternative interpretation, an individual would choose to identify with either her region or her country. Appendix A expands on this argument and shows how our model can easily be reinterpreted as one with endogenous identity.

the share of Canadians that are English-speaking. This is consistent with Li (2010) who in a cross-country study finds that individuals from the ethnic majority group have higher tax morale. What do the different parameters in the utility function capture? A higher value of $\alpha$ implies giving more importance to the public good; a higher value of $\beta$ reduces the concavity of the utility derived from the public good; and a higher value of $\delta$ makes utility more sensitive to small changes in the relative size of the identity group.

## 2.2 Optimal Tax Rate

Public consumption $G(C)$ is financed by a proportional tax rate $\tau(C)$ on income, so $G(C) = \tau(C)Y(C)$. As a result, private consumption $x(\ell, r, C) = (1 - \tau(C))y(r, C)$. Hence, the indirect utility of an individual of identity group $\ell$ who resides in region $r$ and country $C$ is

$$v(y(r,C), S(\ell,C), \tau(C), Y(C)) = y(r,C)(1-\tau(C)) + \alpha S(\ell,C)^{\delta}(\tau(C)Y(C))^{\beta}. \tag{2}$$

The tax rate $\tau(C)$ is decided by majority vote.[8]

**Preferred individual tax rate.** Maximizing (2) with respect to $\tau(C)$ yields the preferred tax rate of an individual of group $\ell$ who resides in region $r$ in country $C$:

$$\tau(\ell, r, C) = \left( \frac{\alpha\beta S(\ell,C)^{\delta} \left(Y(C)\right)^{\beta}}{y(r,C)} \right)^{\frac{1}{1-\beta}}. \tag{3}$$

An individual prefers lower taxes if either her income per capita is higher or the relative size of her identity group is smaller. We therefore say that an individual's preferred tax rate is decreasing in her income-to-identity ratio, denoted by:

$$\tilde{y}(\ell, r, C) = \frac{y(r,C)}{S(\ell,C)^{\delta}} \tag{4}$$

so that

$$\tau(\ell, r, C) = \left( \frac{\alpha\beta (Y(C))^{\beta}}{\tilde{y}(\ell, r, C)} \right)^{\frac{1}{1-\beta}}. \tag{5}$$

[8] Even in non-democratic countries, where people do not vote, regimes are arguably more likely to last if they implement the will of the people. As an example, in African autocracies political power tends to be allocated proportionally to population shares across ethnic group (Francois, Rainer and Trebbi, 2015).

**Country tax rate.** The decision by majority vote then implies that a country's tax rate is decided by the individual with the median income-to-identity ratio, $\tilde{y}_m(C, \delta)$, so

$$\tau(C) = \left( \frac{\alpha\beta(Y(C))^{\beta}}{\tilde{y}_m(C, \delta)} \right)^{\frac{1}{1-\beta}}. \tag{6}$$

The quasi-linear preference structure implies that countries with the same population size spend the same on the public good, independently of their income per capita. This captures the idea that public spending, such as defense and government services, tends to scale up with the size of a country. By the same token, this implies that tax rates are larger in poorer nations than in richer nations. If in a parametrized version of the model this would lead to excessively high taxes in the poorest countries, we could simply assume that in low-income countries individuals attach a smaller weight to the public good. As we will theoretically show later, doing so is irrelevant in the context of our model: reducing the importance of the public good by lowering the value of $\alpha$ in the utility function has no effect on the share of the population that supports secession.

## 2.3 Propensity of Subnational Regions to Secede

To determine the instability of a country, we need to understand whether its residents prefer to remain in the union or secede. If a region $r$ secedes from country $C$ to form its own country, an individual of group $\ell$ in that region faces the same income per capita as before, but a different identity group size, $S(\ell, r)$, a different tax rate, $\tau(r)$, and a different total income, $Y(r)$.[9]

**Region tax rate.** The indirect utility of an individual of group $\ell$ residing in independent region $r$ is

$$v(y(r, r), S(\ell, r), \tau(r), Y(r)) = y(r, r)(1 - \tau(r)) + \alpha S(\ell, r)^{\delta}(\tau(r)Y(r))^{\beta}. \tag{7}$$

The corresponding preferred tax rate of that individual if her region $r$ were independent is then

$$\tau(\ell, r, r) = \left( \frac{\alpha\beta(Y(r))^{\beta}}{\tilde{y}(\ell, r, r)} \right)^{\frac{1}{1-\beta}} \tag{8}$$

[9] In case of secession, a region's income per capita and population might of course vary, but it is unclear in which direction. For example, market fragmentation may lower income per capita, whereas better governance may increase income per capita.

where $\tilde{y}(\ell, r, r) = y(r, r)/S(\ell, r)^{\delta}$. The tax rate of the independent region $r$ is decided by the individual from $r$ with the median income-to-identity ratio, $\tilde{y}_m(r, \delta)$, so

$$\tau(r) = \left( \frac{\alpha\beta(Y(r))^{\beta}}{\tilde{y}_m(r, \delta)} \right)^{\frac{1}{1-\beta}}. \tag{9}$$

Plugging (9) into (7) then gives the utility of a resident of $r$ who belongs to group $\ell$ under independence.

**Preference for secession.** An individual of group $\ell$ residing in region $r$ prefers her region to be an independent country rather than being part of country $C$ if her utility under independence is greater:[10]

$$v(y(r, r), S(\ell, r), \tau(r), Y(r)) > v(y(r, C), S(\ell, C), \tau(C), Y(C)). \tag{10}$$

We can rewrite this condition as

$$y(r) \left( \left( \frac{\beta(Y(C))^{\beta}}{\tilde{y}_m(C,\delta)} \right)^{\frac{1}{1-\beta}} - \left( \frac{\beta(Y(r))^{\beta}}{\tilde{y}_m(r,\delta)} \right)^{\frac{1}{1-\beta}} \right) + S(\ell, r)^{\delta} \left( \frac{\beta(Y(r))}{\tilde{y}_m(r,\delta)} \right)^{\frac{\beta}{1-\beta}} - S(\ell, C)^{\delta} \left( \frac{\beta(Y(C))}{\tilde{y}_m(C,\delta)} \right)^{\frac{\beta}{1-\beta}} > 0. \tag{11}$$

From this condition we can draw three conclusions. First, the benefits from seceding increase in the identity group's regional share, $S(\ell, r)$, and they decrease in the identity group's national share, $S(\ell, C)$. Second, the benefits from seceding increase in a region's income per capita, $y(r)$.[11] Third, whether an individual prefers her region to secede is independent of $\alpha$. This last result is important for the parametrized version of the model in the sense that our findings do not directly depend on whether we take a narrow view of public goods (which would imply a low value of $\alpha$) or a a broad view of public goods (which would imply a high value of $\alpha$). It also means that we could assign a lower value of $\alpha$ to the lowest-income countries, without changing whether an individual supports secession.

Aggregating the preferences for secession across population groups, we can determine for each region $r$ the share of the population that wants to secede, and for each country $C$ the share of the population that prefers to leave the union. This gives us the two measures of instability that we will use in our quantitative exploration of secessionism. One is at the level of subnational regions, and the other at the level of countries.

[10] In our model, an individual's preference for secession is equivalent to her preference for full fiscal decentralization. In that sense, the calibrated version of the model aims to predict the potential demand for secession in the absence of fiscal decentralization. In the empirical validation of the calibrated model we briefly return to this question.

[11] Note that when a region's income per capita increases, the fraction $(Y(r))^{\beta}/\tilde{y}_m(r, \delta)$ decreases. This comparative statics result assumes that the country's income and its median voter do not change.

**Share of region's population that favors secession.** *For region $r$ in country $C$, the share of its population who wants to secede from country $C$ is given by*

$$I(r, C) = \sum_{\ell} \mathbb{I}(\ell, r, C) S(\ell, r) \tag{12}$$

*where*

$$\mathbb{I}(\ell, r, C) = \begin{cases} 1 & : \quad v(y(r,r), S(\ell,r), \tau(r), Y(r)) > v(y(r,C), S(\ell,C), \tau(C), Y(C)) \\ \\ 0 & : \quad \textit{otherwise} \end{cases}$$

**Share of country's population that favors secession.** *For country $C$, the share of its population who wants to secede is given by*

$$I(C) = \sum_{\ell} \sum_{r} \mathbb{I}(\ell, r, C) \frac{N(\ell, r, C)}{N(C)} \tag{13}$$

# 3 Data, Measurement, and Calibration

To explore the quantitative properties of our model, we need data on identity, population and income at the subnational level.

## 3.1 Data and Measurement

**Subnational regions.** Our analysis covers 3,003 first-level administrative regions in 176 countries. To give some examples, first-level administrative regions correspond to states in the U.S., provinces in Canada, regions *(oblasts)* in Ukraine, states in Nigeria, provinces *(províncias)* in Mozambique, regions *(régions)* in France, and states *(Länder)* in Germany. The maps delineating the regions come from GADM version 3.6.

**Identity.** An individual's identity depends on the language she speaks. Although the theory applies more broadly to any dimension of identity, in the empirics we focus on language. We do so partly because of the availability of high-quality language data at the subnational level for the entire globe, and partly because language tends to be the main identity marker of population groups. In principle, we could have considered ethnicity and religion as well. However, in many parts of the world, ethnicity coincides with language, and there are only a handful of regions with secessionist movements that have a distinct religion without also

having a distinct language.[12]

How much utility an individual derives from public consumption depends on the size of her identity group. Depending on whether her subnational region stays within the union or secedes, the relevant size of her identity group is either relative to the country, $S(\ell, C)$, or the region, $S(\ell, r)$. We therefore need to know the share of people who speak each language, not just by country but also by region. To that end, we use spatially disaggregated data on language speakers from a new high-resolution database developed by Desmet, Gomes and Ortuño Ortín (2020). They combine spatial data on population from Landscan with a polygon shapefile of nearly 7,000 languages and a matrix of language speakers by country from the World Language Mapping System, the digitized version of the 17$^{\text{th}}$ edition of the Ethnologue. Using these data, they apply an iterative proportional fitting algorithm to allocate language speakers to 5 km by 5 km grid cells for the entire globe.[13] By aggregating these grid-cell data up to the regional level, we obtain the linguistic composition for all first-level administrative regions.

When measuring an individual's linguistic identity, it is not always clear which linguistic groups to use as primitives. Should Picard and Franco-Provençal, two variations of French, be considered as two different language groups or should they be aggregated into French? More generally, does linguistic identity depend on coarse divisions, such as between Romance and Germanic languages, or on fine divisions, such as between Neapolitan and standard Italian? Since there is no ex-ante good answer to this question, for now we parametrize the coarseness of linguistic divisions, and defer a discussion of its value to the model's calibration. More specifically, we follow Desmet, Ortuño Ortín and Wacziarg (2012) and use the language tree of the Ethnologue to construct language groups at 15 different levels of aggregation,

[12]In Section 6, we describe a novel dataset on secessionist movements around the world. Of the hundreds of subnational regions that have active secessionist movements, only very few have a distinct religion without also having a distinct language: Zanzibar (Tanzania), Northern Ireland (United Kingdom) and Bangsamoro (Philippines). Of those examples, Bangsamoro is not a clear-cut case, because of the existence of different minority languages spoken in the region. Many other examples, where religion may be a driver of secessionism, also have linguistic differences, so they are identified as being distinct by our language data. This is the case of Chechnya (Russia), Tibet (China), Xinjiang (China), Punjab (India), Afar (Ethiopia), Northern Cyprus (Cyprus), Bali (Indonesia), Rohingya (Myanmar) and Nagorno-Karabakh (Azerbaijan). Even Republika Srpska (Bosnia), Sandžak (Serbia) and South Yemen (Yemen) show up as having distinct languages in our dataset, because Bosnian and Serbian are considered different, and the same is true for the various dialects of Arabic spoken in Yemen. Moreover, as we show in Appendix C.2, most coreligionists within a country tend to speak different languages (up to 86), whereas speakers of a language tend to practice the same religion (96% of language groups within countries practice a unique religion). Hence, religious groups tend to be heterogeneous, whereas language groups tend to be homogeneous, suggesting that language may act as a strong group identifier.

[13]Cross-validating the algorithm, they find a correlation of 0.80 when comparing their local diversity measures with the ones in Gershman and Rivera (2018) which are based on census data of subnational regions in sub-Saharan Africa.

$\eta = 1, \cdots, 15$. At the most aggregate level, $\eta = 1$, only the largest language families, such as Indo-European and Niger-Congo, are considered to be different groups, whereas at the most disaggregate level, $\eta = 15$, Picard and Franco-Provençal would be considered distinct from French. This procedure provides us with 15 matrices of the number of language speakers by first-level administrative region, one matrix for each level of linguistic aggregation. An individual's utility then depends on the level of linguistic aggregation. For example, a speaker of Picard would derive more utility from public consumption in France at high aggregation level 1, where Picard belongs to the same linguistic group as French, than at low aggregation level 15, where the two languages belong to different groups.

The relative size of an individual's identity group can also be reinterpreted as the linguistic proximity to her country or region. If we assume that an individual of group $\ell$ has linguistic proximity 1 to anyone speaking her own language and linguistic proximity 0 to anyone speaking a different language, then $S(\ell, C)$ is her average linguistic proximity to the rest of the country and $S(\ell, r)$ is her average linguistic proximity to the rest of the region. Given that linguistic proximity is measured on a 0-1 scale, we can correspondingly define the linguistic distance of an individual to her country and region as, respectively, $1-S(\ell, C)$ and $1-S(\ell, r)$.

In the model, the incentive of a subnational region to secede increases with the linguistic distance of its residents to the country, and it decreases with the linguistic distance of its residents to the region. To visualize these incentives, Figure 1 depicts the population-weighted linguistic distance between region $r$ and country $C$, $\sum_\ell S(\ell, r)(1-S(\ell, C))$. Darker-colored subnational regions are linguistically more distant from the country they belong to, and thus have a stronger incentive to secede. Language groups are based on linguistic aggregation level 15, and thus finely defined. Appendix Figure C.1 shows similar maps for languages defined at a coarser level. In certain regions, such as eastern Africa, southern Europe and northern India, we see a significant decline in linguistic distances of subnational regions to their respective countries when defining languages at this coarser level.

Figure 2 represents the population-weighted linguistic distance between region $r$ and itself, $\sum_\ell S(\ell, r)(1-S(\ell, r))$. Darker-colored subnational regions are linguistically more distant from themselves, and have thus a weaker incentive to secede. Here as well, languages are finely defined, at aggregation level 15. Appendix Figure C.2 shows similar maps for coarser classifications of languages. There, we see important declines in within-region linguistic distances in Europe and parts of Africa. When assessing the identity motive of secessionism, it is important to look at Figures 1 and 2 jointly. For example, while Figure 1 suggests that many regions in sub-Saharan Africa may want to secede because they are linguistically distant from their country, many of those regions also have large linguistic distances to themselves, which lowers their incentive to separate.

Figure 1: Linguistic Distance between Subnational Regions and Countries

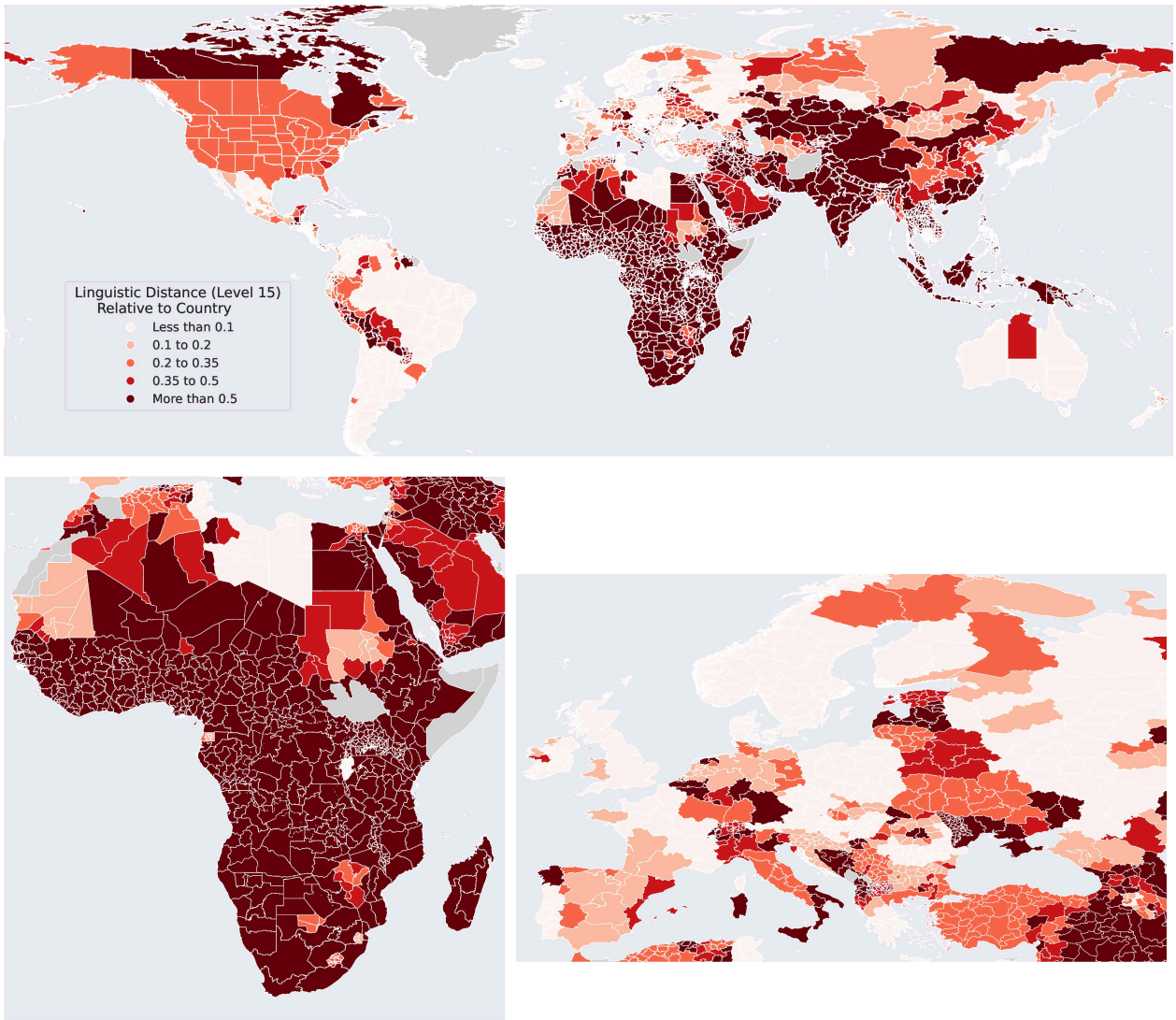



Figure depicts the linguistic distance between each subnational region and the country it belongs to. It is measured as the expected distance between a randomly drawn individual of the subnational region and a randomly drawn individual of the country. Darker colors indicate greater linguistic distances between region and country, and hence a stronger propensity to secede. Languages are defined at its finest level (aggregation level 15).

**Population and income per capita.** Landscan provides estimates for population at a resolution of 30″ by 30″. For the year 2000, we aggregate those data up to first-level administrative regions. As for measuring GDP, we start with data with a resolution of 1° by 1° from the G-Econ 4.0 project at Yale University. For each 1° by 1° grid-cell, we assign GDP to 14,400 smaller 30″ by 30″ grid-cells according to their population weights from Landscan. We then sum those smaller grid-cells to obtain estimates of GDP in the year 2000 for all first-level administrative regions. Figure 3 shows the GDP per capita of subnational regions relative to that of the countries they belong to. Darker-colored regions have higher GDP per

Figure 2: Within-Subnational Region Linguistic Distance

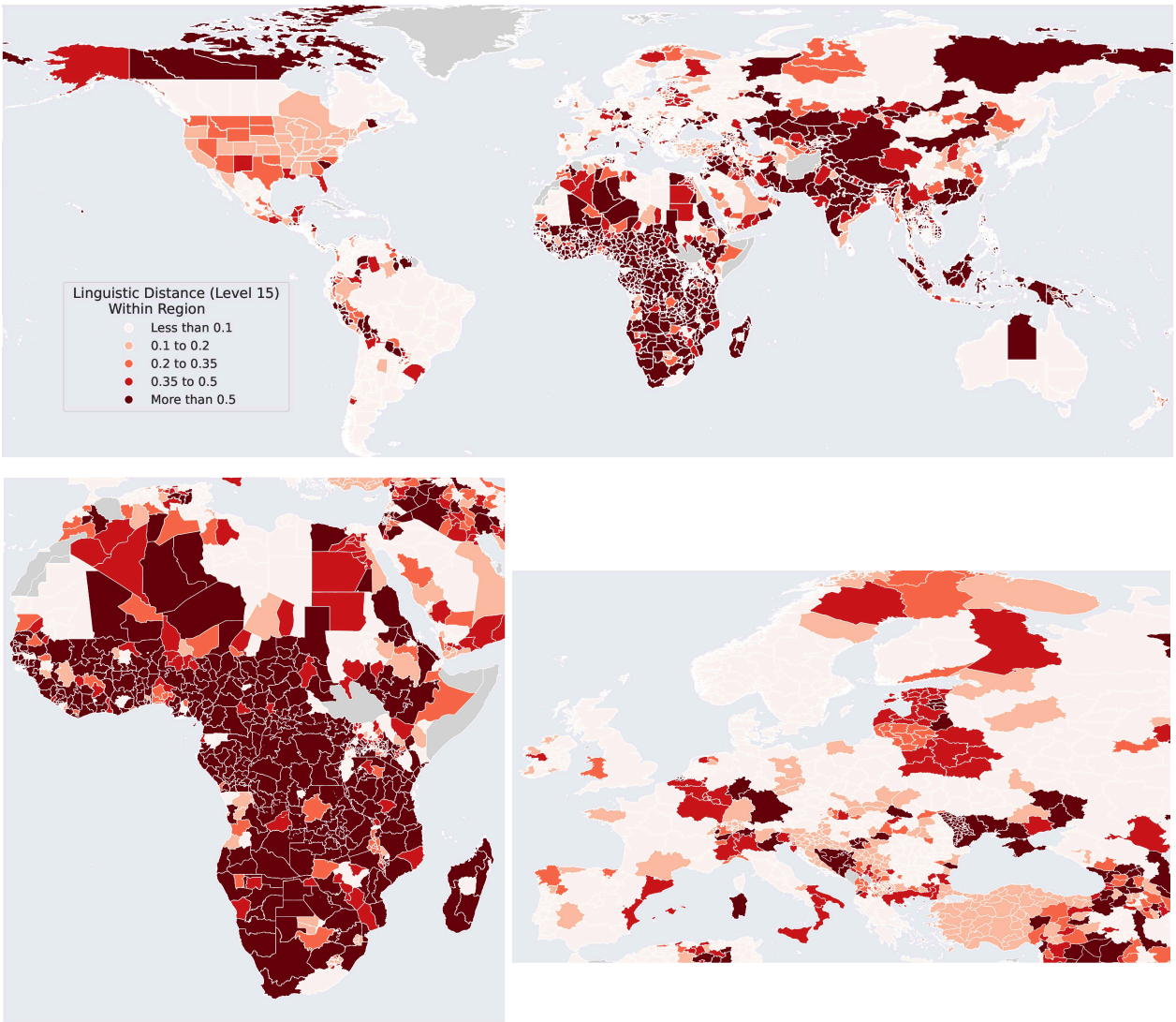



Figure depicts the linguistic distance between each subnational region and itself. It is measured as the expected distance between two randomly drawn individuals of the region. Darker colors indicate a greater within-region distance, and hence a weaker propensity to secede. Languages are defined at its finest level (aggregation level 15).

capita compared to their country, and have thus a stronger incentive to secede.

## 3.2 Calibration

To numerically assess the model's predictions, we need values for $\alpha$ and $\beta$, the parameters that determine the importance and the curvature of the utility derived from public goods, and for $\delta$ and $\eta$, the parameters that determine the sensitivity of utility to the size of the identity group. To choose these parameter values, we consider two calibration strategies: a first calibrates parameters to the current territorial divisions of the world, and a second

Figure 3: Income per Capita Relative to Current Country

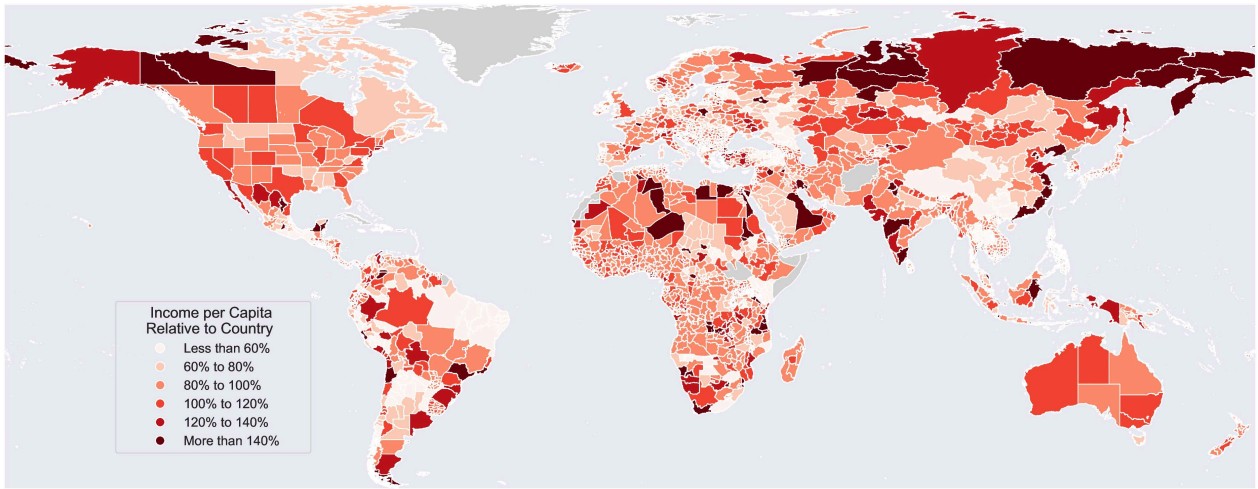



Figure depicts the income per capita of each subnational region relative to that of the country it belongs to in the year 2000. Darker colors indicate a relatively richer subnational region, and hence a stronger propensity to secede.

calibrates to the share of military spending and the regionalist vote share in a subset of countries. As we will discuss, all our key results are robust to whether we use one calibration or the other.

**Baseline calibration.** Our baseline calibration is motivated by the observation that, in spite of widespread territorial tension, the world map is relatively stable. Although there are secessionist movements around the globe, they rarely have the necessary strength to achieve the effective independence of their regions. Since the breakup of the Soviet Union and Yugoslavia, the only successful secessions have been East Timor from Indonesia, Eritrea from Ethiopia, and South Sudan from Sudan. Not only have we seen limited changes to the world map due to secessions of subnational regions, we have also witnessed few modifications due to unions between countries. In the last thirty years, the only internationally recognized unions have been between South and North Yemen and between East and West Germany.

Given that both secessions and unions are rare, we calibrate the parameters to keeping the current world map stable. To do so, we conduct a grid search over the parameter space $\{\alpha, \beta, \delta, \eta\}$ and identify the combination of parameter values that minimizes secessions and unions. More specifically, for each combination $(\alpha, \beta, \delta, \eta)$, we determine the share of subnational regions with a majority in favor of seceding, and the share of neighboring pairs of countries with a majority in both neighbors in favor of uniting. We then choose the parameter values that minimize the average of the model-predicted share of potential secessions and the model-predicted share of potential unions. For the grid search we consider the parameter space $\beta \in [0.05, 0.95]$ in increments of 0.001, $\delta \in [0.05, 0.95]$ in increments of 0.001, and

$\eta \in (1, \ldots, 15)$. Recall that the value of $\alpha$ is immaterial to the support for secessionism, so we set its value to 1. This procedure yields parameter values $\beta = 0.182$, $\delta = 0.198$ and $\eta = 15$.

The reason why we need to use both secessions and unions to calibrate the model is straightforward. If we were to calibrate to only minimizing the number of potential secessions, we would choose to give no importance to identity, and set $\delta = 0$. However, in that case, many neighboring countries would choose to unite. Hence, to get a reasonable value for $\delta$, we need to calibrate to the overall stability of the world map, which requires minimizing both secessions and unions.

**Alternative calibration.** In an alternative calibration, we target the values of $\alpha$ and $\beta$ to data on defense spending, and we target the values of $\delta$ and $\eta$ to the regionalist vote share in a subgroup of advanced democracies. More specifically, for any given combination of $\delta$ and $\eta$, we choose the values of $\alpha$ and $\beta$ that minimize the sum of squared errors between model-predicted government spending and observed average defense spending in developed countries between 1995 and 2015 according to the *Government Finance Statistics Yearbook* of the International Monetary Fund. We then choose the values of $\delta$ and $\eta$ that best predict the secessionist vote share in the three countries with the highest regionalist vote share between 1995 and 2014 according to Sorens (2008). This calibration yields parameter values $\alpha = 8.0902$, $\beta = 0.1264$, $\delta = 0.16$ and $\eta = 9$.

When targeting public spending, we focus on defense, because it comes close to what might be considered a country-level public good. One possible concern is that secessionist conflict may drive defense spending. To avoid this concern, we only consider developed countries with a GDP per capita of at least 50% of that of the European Union. Another possible concern is that we are not giving a big enough role to public goods, because defense spending typically does not make up more than 1-2% of GDP. However, if we were to multiply defense spending by, say, a factor of 10, this would simply translate into a different calibrated $\alpha$. As we have shown in (11), this would leave the population share in favor of secession unchanged. One might also be concerned that calibrating to developed countries in the context of quasi-linear preferences implies excessive predicted public spending in developing countries. While this could easily be addressed by setting $\alpha$ to a lower value in less developed countries, there is no need to do so, since the results are invariant to changes in $\alpha$.

# 4 Predicted Propensity to Secede

This section explores the predictions of our calibrated model for the instability of countries.

**Baseline calibration.** Using our baseline calibration, Figure 4 shows the model-based share of each region's population in favor of secession. The top 10% of regions with strongest support for secessionism include regions such as Tibet (China), Southern Nations (Ethiopia), Bavaria and Saarland (Germany), Aceh (Indonesia), Lombardia and Sardinia (Italy), Okinawa (Japan), Friesland (Netherlands), Arad (Romania), Tatarstan (Russia), Western Cape (South Africa), and Galicia and Catalonia (Spain).[14] Some of these regions have witnessed violent territorial conflict (Aceh and Tibet), some feature significant pro-independence movements (Catalonia and Okinawa), some have distinct regional cultures (Friesland, Galicia and Sardinia), and others are border regions that in the past were part of a neighboring country (Saarland).

For the same baseline calibration, Figure 5 depicts the model-based share of each country's population that prefers their subnational region to secede. Ignoring small island nations, the most unstable countries for each of the six major continents are India, Italy, Papua New Guinea, South Sudan, Guatemala, and Peru. When comparing secessionism by the world's main regions as classified by the World Bank, the highest support in terms of population share is found in South Asia (14%), followed by sub-Saharan Africa (13%). The lowest support is found in North America, where less than 1% of the population favors secession. Appendix Table C.1 gives the full list of countries, and provides additional measures of instability, such as the share of regions with a majority in favor of secession and the share of the population residing in regions with a majority in favor of secession.

**Alternative calibration.** Turning to the alternative calibration, Appendix Figure C.4 displays the regional and country population shares in favor of secession, and Appendix Table C.2 gives the full list of countries with their respective instability measures. The most unstable countries for each of the six continents are Pakistan, Belgium, Papua New Guinea, Ethiopia, Canada and Bolivia. The correlation between the country population shares in favor of secession in the baseline calibration and the alternative calibration is 0.59. One difference is that there is a slightly smaller average country population share in favor of secession (6.2%, compared to 6.9% in the baseline calibration).[15] One reason is that linguistic groups are measured at a coarser level ($\eta$ is 9, compared to 15 in the baseline).

[14] Our analysis focuses on the propensity to secede of each subnational region separately. Of course, our approach also allows for estimating the propensity to secede of coalitions of subnational regions. However, this would require, on the one hand, recalibrating the model, and on the other hand, taking a stance on what those relevant coalitions might be. We leave this question for future research.

[15] This may come as a surprise, since the baseline calibration targets the stability of the current world map. However, in that baseline calibration, we minimize not just the share of regions with a majority in favor of secession, but also the share of neighboring country pairs with majorities in favor of uniting.

Figure 4: Model-Based Share of Regional Population in Favor of Secession

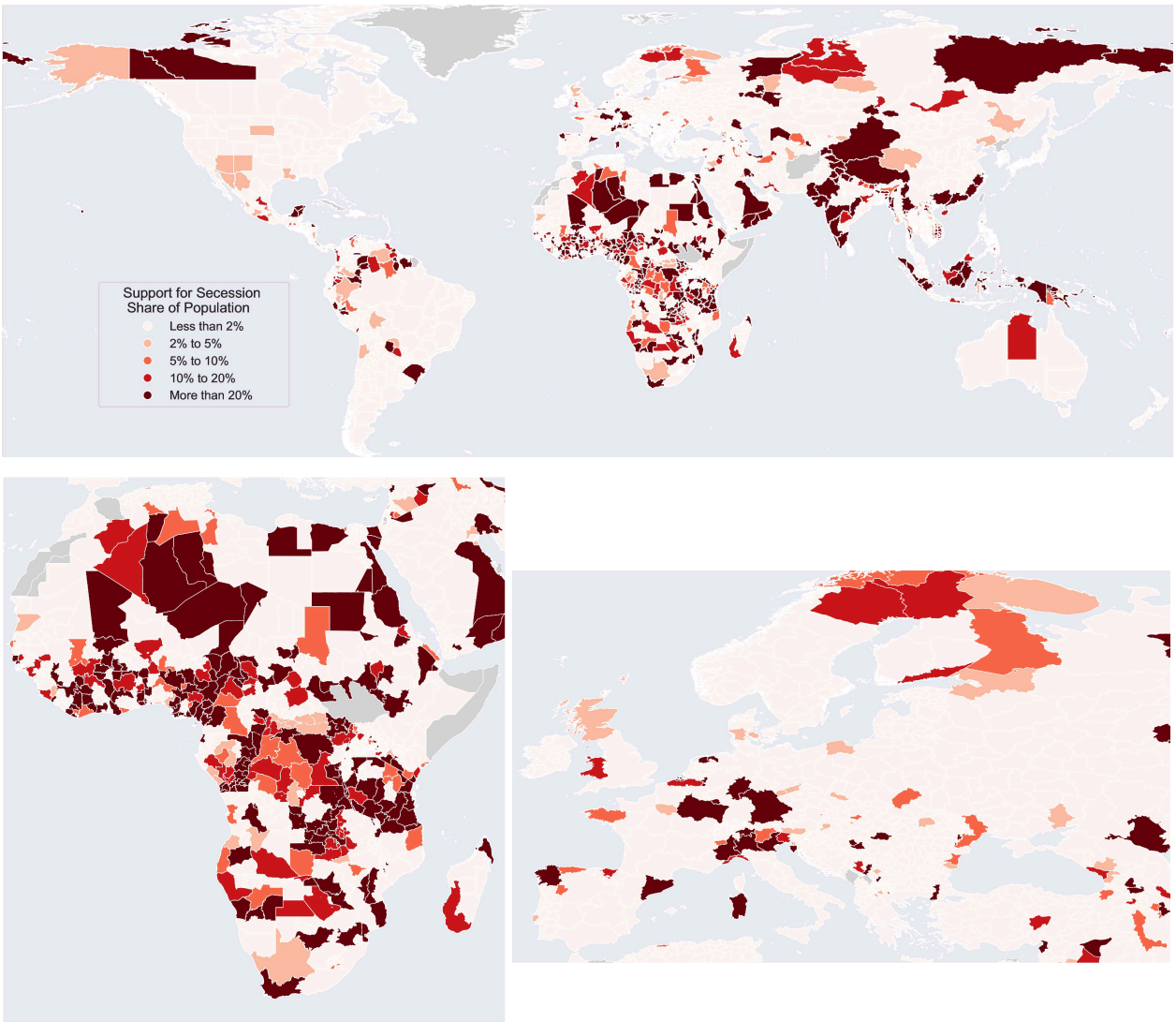



Figure depicts the model-generated share of the population in favor of secession for each of the 3,003 subnational regions. Results are based on the baseline calibration.

# 5 Is Secessionism Mostly about Identity or Income?

In the model, both identity and income per capita determine the incentives to secede. In this subsection we analyze their relative importance in the calibrated version of the model.

**Identity and income per capita differences.** We start by conducting two counterfactual exercises. In a first exercise, we remove differences in subnational income per capita by assuming that each region's income per capita is equal to the country average. In that case, the incentives to separate depend on linguistic identity, and not on income. In a second exer-

Figure 5: Model-Based Share of Country Population in Favor of Secession

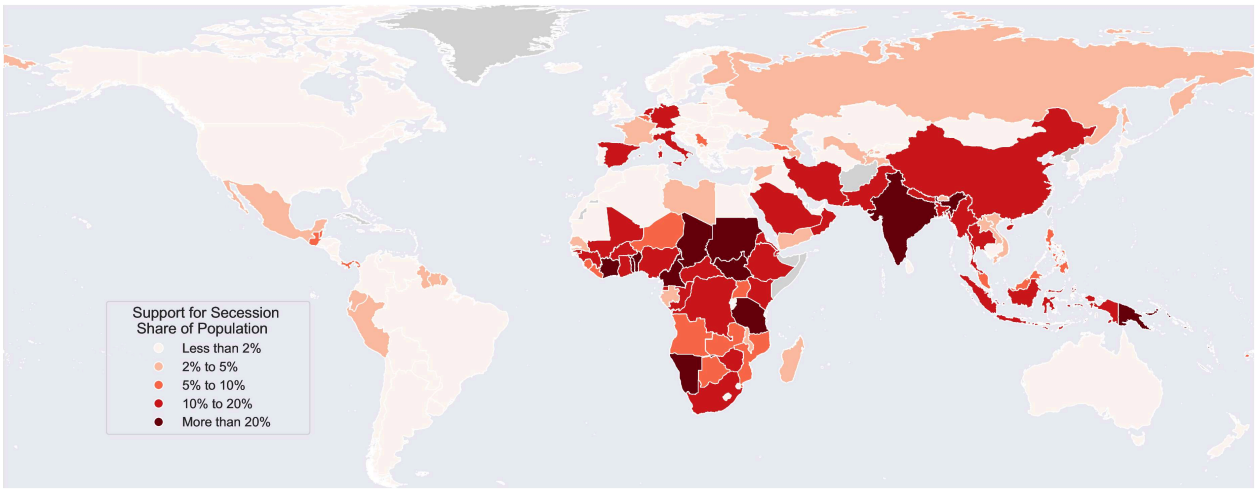



Figure depicts the model-generated share of the population in favor of secession for each country. Results are based on the baseline calibration.

cise, we remove identity differences by assuming that everyone speaks the same language.[16] In that case, the incentives to secede depends on income per capita, and not on linguistic identity. Appendix Table C.3 reports the results of these counterfactual exercises. More specifically, it gives the change in a country's population share in favor of secession when either only identity matters or only income per capita matters.

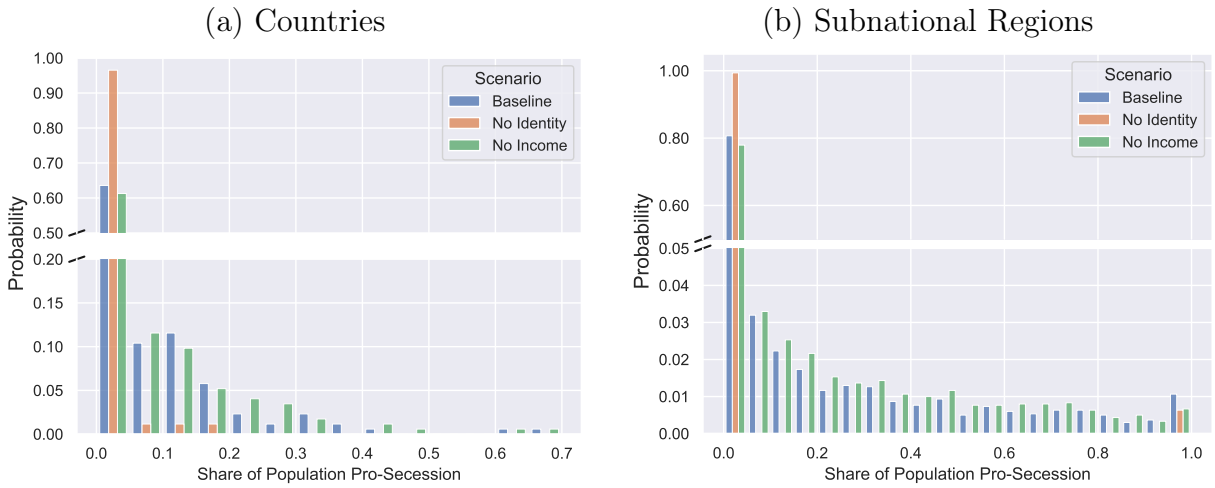
When removing differences in income per capita and maintaining linguistic identity as the main determinant, the average support for secession does not drop. In fact, it even increases, by 0.6 percentage points at the country level (from a baseline of 6.9%) and by 0.9 percentage point at the regional level (from a baseline of 7.5%). This may come as a surprise, since secessionism tends to be stronger in regions with higher income per capita. Hence, we would expect that equalizing income per capita should on average weaken the propensity to secede. However, there are also many subnational regions with lower income per capita and a distinct linguistic identity. In those regions, equalizing income per capita tends to strengthen the support for independence. Quebec provides a good example. There, setting income per capita to the Canadian average leads to an increase in secessionist sentiment. It turns out that examples such as Quebec dominate, so that eliminating within-country

[16] We consider two different methods of eliminating identity differences in the context of our model. A first sets $\delta$ to 0, and hence $S(\ell, C)^{\delta}$ to 1, in equation (1). In that case, the contribution of the public good to the utility of an individual of group $\ell$ residing in region $r$ of country $C$ is $\alpha G(C)^{\beta}$. A second sets $S(\ell, C)^{\delta}$ to its population-weighted country average $\sum_{\ell} S(\ell, C)^{1+\delta}$. In that case, the contribution of the public good to the utility of an individual of group $\ell$ residing in region $r$ of country $C$ is $\alpha \left(\sum_{\ell} S(\ell, C)^{1+\delta}\right) G(C)^{\beta}$. With both methods, the utility of the public good no longer depends on the linguistic group one belongs to. Below we report results based on the first method, but the results are qualitatively unchanged when using the second method.

income per capita differences actually slightly increases the average support for secession.

In contrast, when removing linguistic identity differences and maintaining differences in income per capita, the support for secession drops on average by 6.4 percentage points at the country level (from a baseline of 6.9%) and by 6.9 percentage points at the regional level (from a baseline of 7.5%). This implies that in many countries secessionism loses all its support. Again, there are a few exceptions: in Namibia 14.5% of the population continues to support secession (down from 20.2% in the baseline) and in Sudan support is at 17.2% (down from 22.4% in the baseline).

Figure 6: Predicted Share of Population in Favor of Secession




Graph shows the distribution of the population share in favor of secession for (a) countries and (b) subnational regions, for three different scenarios: the baseline where both identity and income play a role, the scenario without identity differences, and the scenarios without income per capita differences.

Figure 6 summarizes our findings by showing the distribution of the population share in favor of secession for the baseline and the two counterfactuals. Panel (a) depicts the distribution at the level of countries, whereas Panel (b) shows the same information at the level of regions. Removing income per capita differences between subnational regions has almost no effect on the distribution of support for secession, whereas eliminating identity differences causes an important leftward shift of the distribution toward zero. More specifically, in the baseline 41% of countries have less than 1% of their population favoring secession, compared to an almost identical 39% in the absence of income differences, but a much higher 95% when countries become linguistically homogeneous. At the level of subnational regions, the results are equally stark. In the baseline, 47% of subnational regions have zero support for independence, compared to a similar 41% when per capita income is equalized across regions. However, if everyone were to speak the same language, more than 99% of regions would no longer have any separatist support. This result seems to vindicate the old adage of "one

nation, one language".

Taken together, these two counterfactual exercises suggest that linguistic identity is much more important than income per capita for secessionism. One potential issue with this conclusion is that the unimportance of income per capita might be due to averaging across subnational regions. For example, when we equalize income across regions in South Africa, secessionism collapses in the relatively rich Western Cape, whereas it emerges in the relatively poor Eastern Cape. To investigate whether the absence of income per capita as a significant driver of secessionism is due to averaging, we limit our attention to the subgroup of regions with a strictly positive share of the population in favor of secession in the baseline. When removing income per capita differences, the support for secession in this subgroup of regions does drop, but not by much, from a baseline average of 44% of the population to 38%. This confirms that secessionism is not very sensitive to changes in income per capita.

These findings are confirmed when using the alternative calibration. In that case, when removing within-country differences in income per capita, the average support for secession hardly changes, from 6.4% of a region's population to 6.3%. In contrast, when removing within-country identity differences, the average support for secession collapses to 0.5%. Once again, we might worry that the weak role of income per capita is due to averaging across subnational regions. This is not the case: when only considering the subgroup of regions with strictly positive support for secession in the baseline, we find that secessionism falls only slightly, from an average of 47% of a region's population to 41%.

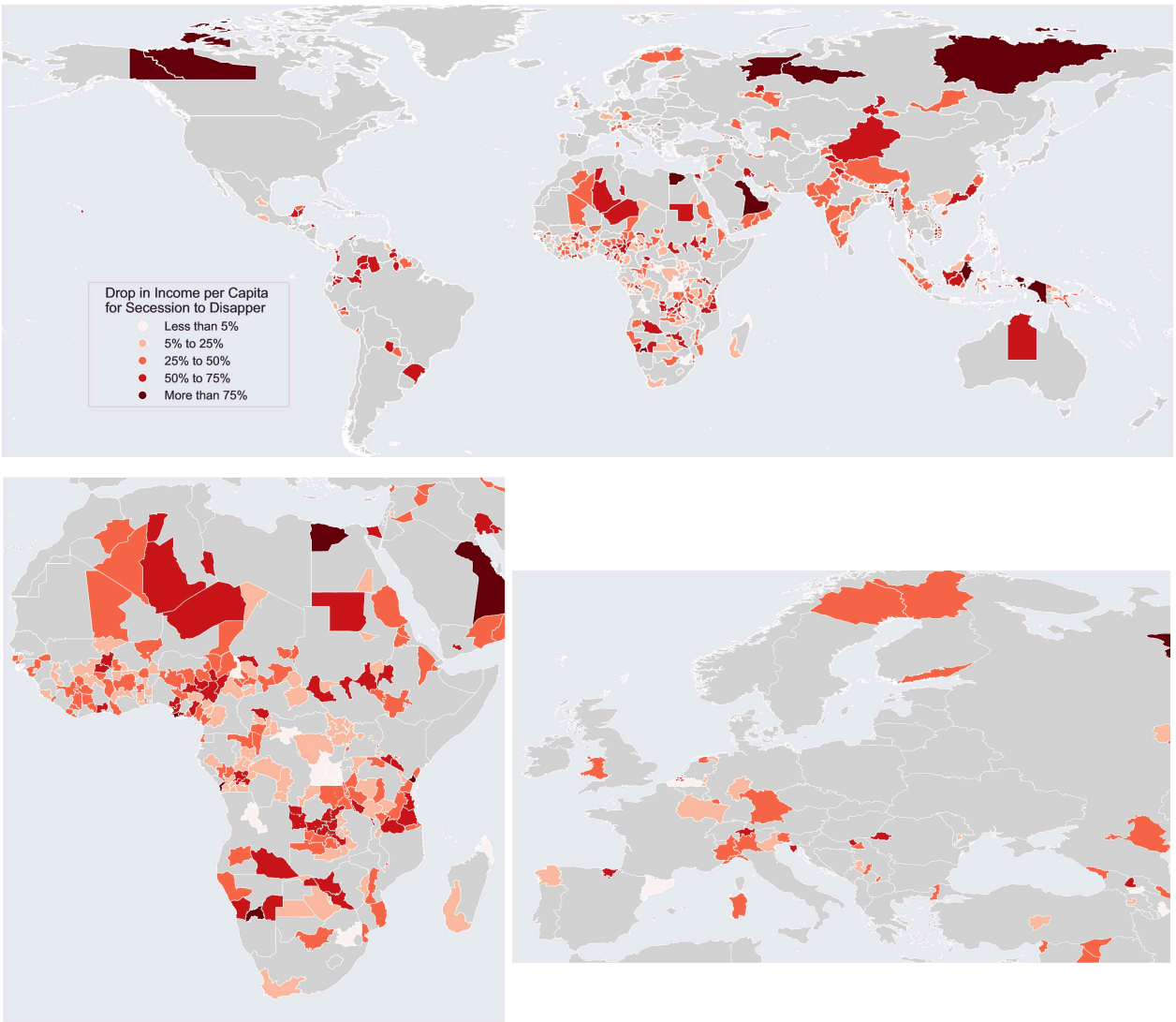
All of this suggests that identity is the essential driver of secessionism, with economic forces playing a much smaller role. Next, we conduct further counterfactual analysis to deepen our understanding of the relative importance of the different drivers of secessionism.

**Sensitivity of secessionism to changes in income per capita.** In the baseline calibration, subnational regions with at least 10% of their population in favor of secession are on average 13% richer than the countries they belong to. Our counterfactual exercises so far have shown that removing this income advantage does not suffice for secessionism to weaken. Might this be because setting their income per capita to the national average is only a small shock? How far should income per capita fall for those subnational regions to give up on secessionism?

To answer this question, we look at each subnational region with at least 10% of its population in favor of secession, and compute by how much income per capita needs to decrease for secessionist support to fall below 1%.[17] The maps in Figure 7 depict our findings. In Catalonia (Spain), a drop in income per capita by 5% is enough for secessionism to vanish.

[17] Of course, when a region's income per capita declines, it also shrinks the country's overall economy.

Figure 7: Drop in Income per Capita for Secessionism to Disappear




For each subnational region with more than 10% of its population in favor of secession, Figure depicts the necessary drop in income per capita for the support of secession to drop below 1%.

For most of the other examples we discussed before, the corresponding figures tend to be larger: Tibet (China), -30%; Southern Nations (Ethiopia), -40%; Saarland (Germany), -35%; Aceh (Indonesia), -35%; Lombardia (Italy), -35%; Sardinia (Italy), -40%; Okinawa (Japan), -30%; Tatarstan (Russia), -30%; Western Cape (South Africa), -20%.

Going beyond specific examples, income per capita needs to drop on average 43% for secessionist support to vanish in subnational regions with at least 10% of its population in favor of secession. Using the alternative calibration, the corresponding drop in income per capita is 44%. Where does this leave us in terms of assessing the importance of economic forces? On the one hand, income per capita matters: poor regions have less incentives to

separate. On the other hand, subnational regions need to be economically far behind for secessionism to substantially weaken. Therefore, secessionism is only weakly sensitive to changes in income per capita. This confirms our previous conclusion.

**Sensitivity of secessionism to within-region identity heterogeneity.** While identity differences with the rest of the nation strengthen secessionist tendencies, identity differences within subnational regions diminish secessionist tendencies. Indeed, regions have less reason to become independent if their own linguistic identity is diverse. To further assess the importance of identity for secessionism, we focus on this within-region identity heterogeneity. More specifically, we explore what happens if individuals ignore within-region linguistic identity differences in the case of independence. In this counterfactual exercise, we do not change the linguistic composition of subnational regions. Instead, we simply assume that individuals cease to care about within-region identity differences in the case of secession.[18] This amounts to setting $\delta$ equal to zero if a region chooses to become independent, while maintaining the original parameter value if a region remains part of the union.

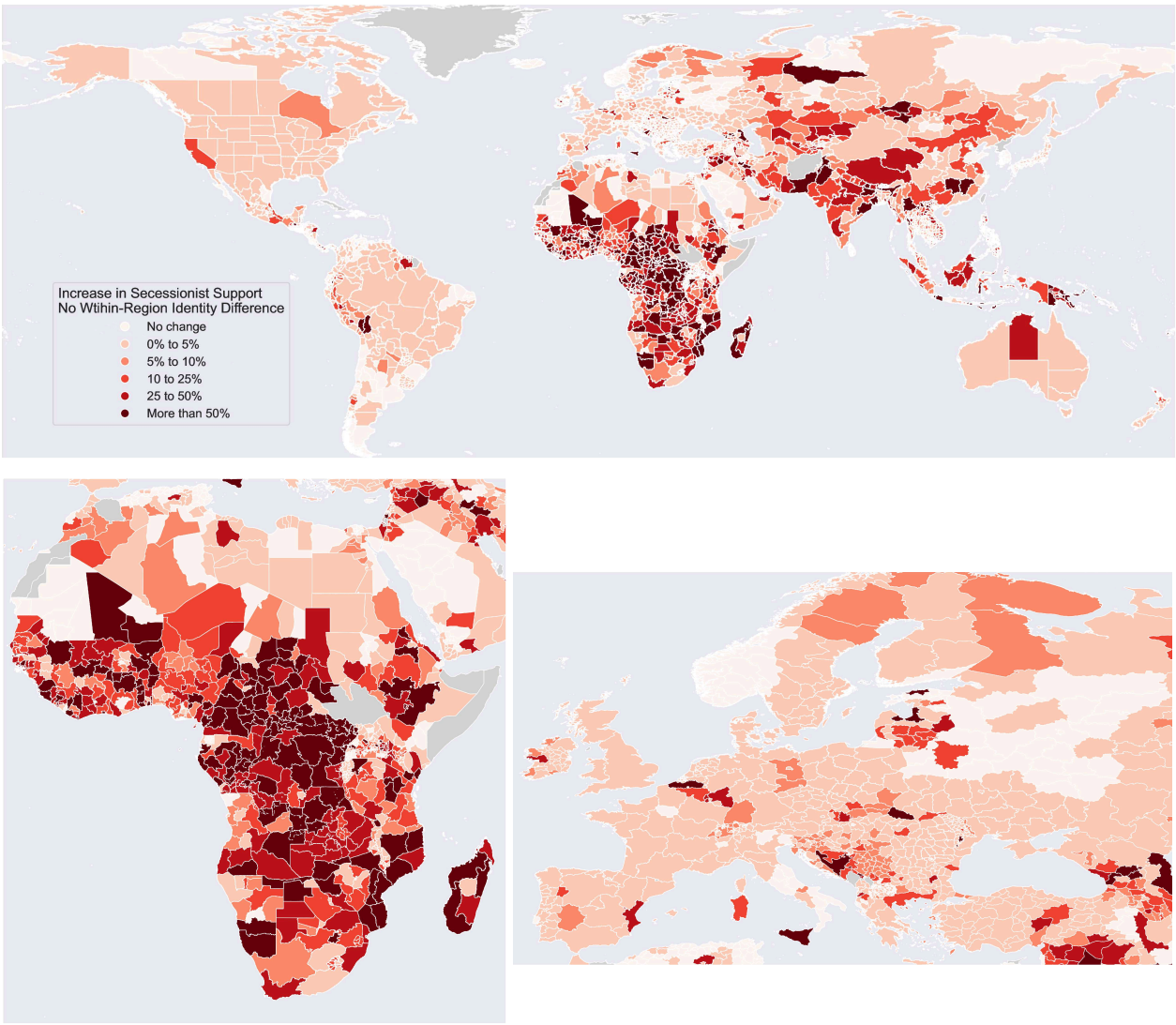
The maps in Figure 8 depict the percentage point increase in support for secessionism if individuals ignore within-region identity heterogeneity in the case of independence. We observe a large increase in support for secessionism. At the country level, we witness a more than tripling of the support for regional separation: from 6.9% in the baseline to 24.0% in this counterfactual exercise. And at the regional level, the support increases from 7.5% to 20.0%. Under the alternative calibration, the relative increase at the country level is similar: from 6.2% in the baseline to 18.2%. The high degree of sensitivity of secessionism to within-region linguistic identity differences confirms our previous conclusion: identity is a key driver of separatist sentiment.

**Secessionism and policy.** Our counterfactual exercises strongly suggest that identity, more than economics, is key to understanding secessionist tendencies. Because the demand for separatism is not very sensitive to income, economic policy is probably of limited help to stave off secessionist threats. For example, we have shown that setting regional income per capita to the national average does not change, or can even worsen, the overall level of territorial instability.

Instead, creating a common identity is more likely to weaken the propensity to secede. This is reminiscent of the nation-building efforts of the 19$^{\text{th}}$ century. Alesina, Giuliano and Reich (2021) describe how the introduction of a "national language", often through com-

[18] An alternative exercise would be to have everyone in a region adopt a common language. However, in that case the results would depend on which language becomes the common language.

Figure 8: Increase in Secessionist Support when Eliminating Within-Region Identity Differences




For each subnational region, Figure depicts the increase in support for secession if people were to ignore within-region identity differences in the case of independence.

pulsory education, is central to nation-building efforts. For example, at the time of Italian unification, at most 10% of its population spoke Italian. Linguistic homogenization was seen as key to keep the newly-created country united. Likewise, the expansion of the Russian Empire during the late-nineteenth century was accompanied by Russification through education. France went through a similar process of nation-building. As Weber (1979) argues, the village school was the "ultimate acculturation process that made the French people French". Of course, the relevant question is to what extent the homogenization of language use weakens secessionist tendencies. Recent work by Blanc and Kubo (2021) shows that

French municipalities that benefited more from state-sponsored education in the nineteenth century displayed greater participation in the Resistance during WWII and received more votes against the 1969 Referendum on Regionalization.

Of course, in the 21st century, policies that promote the use of a common language are often politically and socially no longer acceptable. Instead, countries might still attempt to instill a sense of nationhood and a common identity through other means. Rohner and Zhuravskaya (2023) provide an overview of different approaches to nation building.

# 6 Assessing Model Performance

The credibility of our counterfactual analysis depends on the credibility of our calibrated model. To assess how well the model performs, we correlate the model-predicted measures of instability to actual measures of instability, and we explore whether the calibrated model can account for the breakup of the Soviet Union.

## 6.1 Correlation with Secessionist Movements

This section correlates our model-based measures of the demand for secession to actual secessionist activity.

**A novel database of secessionist movements.** The most direct way of assessing the accuracy of our quantitative model is to compare the propensity to secede as predicted by the model to actual secessionist activity around the world. To that end, we compile a novel global dataset on the presence, number, intensity, and geographic distribution of secessionist movements. Unlike previous efforts, our dataset provides information on secessionist movements at the level of subnational regions and it focuses on active secessionist movements.[19]

Our starting point is a global list of 2,529 active secessionist movements recorded by Wikipedia.[20] These secessionist movements include political parties, militant and civil organizations, and social and ethnic movements. Some are armed and violent; others are non-violent and unarmed. As an example, for the case of Catalonia in Spain, the list includes political parties, such as Esquerra Republicana de Catalunya and Junts per Catalunya, civil organizations, such as Assemblea Nacional Catalana and Òmnium Cultural, and youth

[19]For example, Sambanis, Germann and Schädel (2018) provide comprehensive data at the level of countries. Another difference with our data is that they include certain ethnic groups that do not seek secession, but rather regional autonomy or indigenous land rights like the Cacarica or Paez communities in Colombia or the Hazaras in Afghanistan.

[20]The list of active secessionist movements comes from Wikipedia and was downloaded on October 12, 2020. It is organized by continent (Africa, Asia, Europe, North America, Oceania, South America).

organization, such as Arran and La Forja. As another example, for the case of Balochistan in Pakistan, it includes political parties, such as the Baloch National Movement, and militant organizations, such as the Baloch Liberation Army.

Using this global list, we link each secessionist movement to one or more first-level administrative regions. This determines for each subnational region in our dataset whether there exists any active secessionist movement. We find that secessionist movements are active in 511 first-level administrative regions in 94 countries. To get a measure of the intensity of secessionist activity by region, we use two indicators. The first is simply the number of secessionist movements by region. The second is the number of visits to the Wikipedia webpages of these secessionist movements. For this, we sum the page views over all of Wikipedia's language versions for the period 2015-2020. Using the example of Catalonia, the political party Esquerra Republicana de Catalunya received 2,037,951 views between 2015 and 2020, whereas the civil organization Assemblea Nacional Catalana had 1,212,116 visits and La Forja got only 22,838 page views.

One potential concern with our dataset is whether it covers all regions of the world equally, and more particularly, whether there is a bias towards developed countries. We will address this concern in the robustness discussion at the end of this subsection.

**Cross-country variation in secessionist activity.** We now explore how predictive our model-based measures of instability are of actual secessionist activity, both across and within countries. Starting with a cross-country approach, we use the following empirical specification:

$$S_c = \beta \hat{S}_c + \delta N_c + \sum_j \gamma_{cj} + \varepsilon_c, \tag{14}$$

where $S_c$ is one of our measures of secessionist activity at the country-level, $\hat{S}_c$ is one of our country-level model-based measures of demand for secession, $N_c$ is the number of regions in the country, $\gamma_{cj}$ is a complete set of World Bank regional fixed effects, and $\varepsilon_c$ is a random error. Panel A of Table 1 reports our findings for these country-level regressions of actual secessionist activity on model-based measures of secessionism, controlling for World Bank region fixed effects and the number of subnational regions per country. In column (1) we see that the model-based number of secessionist regions by country is strongly predictive of the actual number of secessionist regions.[21] Columns (2) and (3) show that the same is true for the model-based share of the population that favors separatism and the model-based measure of at least one region being in favor of secession. These different model-based

[21] To be a secessionist region, the model requires 50% of its population to favor independence whereas the data require there to be at least one secessionist movement.

Table 1: Predicted Secessionism and Contemporary Secessionist Activity

| Panel A: Country-Level Analysis | Interest in Secession | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Secessionist Activity in Country | | | | | | Number of Wikipedia Page Views (IHS) | | |
| | # Secessionist Regions (IHS) | | | # Secessionist Groups (IHS) | | | Secessionist Organization | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| # Secessionist Regions (IHS) | 0.344*** | | | 0.542*** | | | 1.751** | | |
| | (0.110) | | | (0.185) | | | (0.730) | | |
| Share of Pop. Pro-Secession (IHS) | | 2.346*** | | | 3.744** | | | 12.112* | |
| | | (0.888) | | | (1.534) | | | (6.316) | |
| At least 1 Region Pro-Secession | | | 0.605*** | | | 0.990*** | | | 2.999** |
| | | | (0.179) | | | (0.291) | | | (1.164) |
| # Regions | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.31 | 0.31 | 0.32 | 0.26 | 0.25 | 0.27 | 0.18 | 0.17 | 0.18 |
| Observations | 173 | 173 | 173 | 173 | 173 | 173 | 173 | 173 | 173 |

| Panel B: Regional-Level Analysis | Interest in Secession | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Secessionist Activity in Region | | | | | | Total Number of Wikipedia Page Views (IHS) | | |
| | Secessionist Region | | | # Secessionist Groups in Region (IHS) | | | Secessionist Organization | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Sh. Pop. Pro-Secession (IHS) | 0.238*** | | | 0.459*** | | | 1.964*** | | |
| | (0.057) | | | (0.127) | | | (0.659) | | |
| At least 10%+ Pro-Secession | | 0.125*** | | | 0.239*** | | | 1.094*** | |
| | | (0.028) | | | (0.062) | | | (0.346) | |
| At least 50% Pro-Secession | | | 0.133*** | | | 0.280*** | | | 1.163*** |
| | | | (0.038) | | | (0.100) | | | (0.439) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.21 | 0.21 | 0.21 | 0.24 | 0.24 | 0.24 | 0.21 | 0.22 | 0.21 |
| Observations | 3003 | 3003 | 3003 | 3003 | 3003 | 3003 | 3003 | 3003 | 3003 |

Notes: Table reports results from regressions of different measures of secessionist activity from Wikipedia on model-generated measures of the demand for secession. Panel A does this at the level of countries, and includes World Bank region fixed effects and the number of subnational regions as additional controls, whereas Panel B does this at the level of subnational regions and includes country fixed effects. IHS denotes variables have been transformed using the inverse hyperbolic sine transformation. Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

measures of secessionism are also statistically significant predictors of the actual number of secessionist movements (columns (4) to (6)) and the number of Wikipedia page views of secessionist organizations (columns (7) to (10)). The magnitudes of the effects are large: countries that according to the model have at least one secessionist region have 99% more

secessionist movements in the data and 300% more page views of the Wikipedia entries of these movements. By including regional fixed effects in our estimating equation, we are not allowing the model-based measures of secessionism to reflect regional differences in actual secessionist activity. When dropping regional fixed effects, Appendix Table C.4 shows that our model-based measures continue to be highly predictive of actual secessionist activity.

**Within-country variation in secessionist activity.** Next, we estimate the association between our model-based measures of demand for secession and secessionist activity within countries using the following empirical specification:

$$S_{rc} = \beta \hat{S}_{rc} + \sum_{c} \gamma_{rc} + \varepsilon_{rc}, \tag{15}$$

where $S_{cr}$ is one of our measures of secessionist activity in region $r$ of country $c$, $\hat{S}_{rc}$ is one of our regional-level model-based measures of demand for secession in region $r$ of country $c$, $\gamma_{rc}$ is a complete set of country fixed effects, and $\varepsilon_{rc}$ is a random error. We cluster standard errors at the country-level. Panel B in Table 1 reports our findings for these subnational regressions of actual regional secessionist activity on our model-based regional measures of secessionism. Column (1) shows that the model-based share of the region's population in favor of secession is highly predictive of the region having at least one secessionist movement in reality. Columns (2) and (3) show that the same holds for model-based indicator variables that measure whether a region has at least 10% or 50% of its population that favors independence. These different model-based measures of secessionism are also highly predictive of intensive measures of secessionist activity across regions, such as the number of secessionist groups (columns (4) to (6)) and the number of Wikipedia page views of secessionist movements in the region (columns (7) to (10)). The magnitudes of the effects are substantial: having at least 10% of a subnational region's population in favor of secession according to the model is associated with a 24% increase in the number of secessionist groups in that region and 109% more visits to the Wikipedia entries of those groups. These results control for country fixed effects, and hence account for any unobserved heterogeneity at the country level, such as national institutions and culture. Overall, these large, positive, and statistically significant associations suggest that our model-based measures of potential demand for secession at the country and regional levels capture some of the forces underlying actual secessionist movements around the world.[22]

[22]The results in Table 1 use the inverse hyperbolic sine transformation, which is unbiased in its treatment of zeros. Appendix Section C.7 reruns the entire analysis using logarithms to transform the measures. Reassuringly, the results are qualitatively similar.

**Robustness.** One potential concern is whether our Wikipedia data have equal coverage across the globe. We address this concern in different ways. We start by replicating our analysis dropping World Bank regions one at a time, obtaining similarly significant associations (Table C.5). We then compare our database to the one on self-determination movements by Sambanis, Germann and Schädel (2018). This alternative database covers 120 countries for the period 1945-2012, and it expands on previous data on self-determination movements, such as the Minorities at Risk (MAR) project (Gurr 1993, 2000), and the Peace and Conflict reports from the Center of International Development and Conflict Management (Marshall and Gurr, 2003, 2005). While it provides a comprehensive list of self-determination movements, their data are not linked to subnational regions, and they do not include a measure of the intensity of the different movements.

To further explore the quality of our data and the robustness of our results, we use the Sambanis et al. (2018) database in two ways. First, we compare, at the level of countries, the number of secessionist groups from Wikipedia with the number of self-determination movements during the 2000-2012 period from Sambanis et al. (2018), finding a correlation of 0.71 ($p < 0.001$). Second, Appendix Table C.6 Panel A reports the association between our model-based measures of instability and the presence and number of self-determination movements during the 2000-2012 period according to Sambanis et al. (2018). The results show that our model-based measures are strong predictors of the presence of self-determination movements (columns 1-3) as well as their number (columns 4-6), after accounting for World Bank region fixed effects and the number of regions in a country. Panel B shows results when we combine our data from Wikipedia with those of Sambanis et al. (2018). Specifically, we code a country as having separatist movements if it has at least one according to one of these sources. We also use the maximum number of movements across both data sources as a proxy for the actual number of movements in the country. The results are unchanged.

## 6.2 Correlation with Other Measures of Instability

This section correlates our model-based measures of instability to state fragility, regional autonomy, and conflict.

**State fragility.** Another expression of country instability is its vulnerability to collapse. The Fragile States Index (FSI), developed by the Fund for Peace, uses a conflict assessment framework to evaluate this type of vulnerability.[23] The FSI is a composite measure covering

[23]The data is accessible at https://fragilestatesindex.org/. Accessed on June 23, 2021.

various aspects of state power and fragility.[24] Panel A of Table 2 reports the association between our model-based measures of secessionism and a country's ranking (columns (1) to (3)) and value according to the FSI (columns (4) to (6)). In all cases but one, the association is statistically significant at the 1% level. To give a sense of the magnitude of the effects, a country with at least one region in favor of secession changes its rank by 24 positions in the direction of greater fragility compared to one that has no secessionist regions. Such a country also increases its FSI value by an average of 10 points, a large change close to half a standard deviation of FSI. Appendix Table C.7 shows that the positive association holds for all 12 subcomponents, and the association is statistically significant at the 10% for all but three subcomponents.

**Regional autonomy.** In many countries, the demand for secessionism expresses itself in greater regional autonomy, rather than in an outright push for independence. In fact, in our model a region that becomes an independent country is equivalent to a region that enjoys full fiscal decentralization. We therefore look at the association between our model-based measures of secessionism and the Regional Authority index (RAI), developed by Hooghe et al. (2016).[25] The index measures the authority of subnational governments in 95 democracies or quasi-democracies on an annual basis from 1950 to 2018. It captures different aspects of self-rule and shared-rule, based on ten subcomponents related to institutional depth, policy scope, fiscal autonomy, borrowing autonomy, representation, law making, executive control, fiscal control, borrowing control, and constitutional reform. The last 3 columns of Panel A in Table 2 report the association between our model-based measures of secession and the Regional Autonomy Index (RAI). Once again, the estimated associations are positive and statistically significant. For example, a country with at least one secessionist region has an RAI that is 7.9 points higher (equivalent to 0.8 of a standard deviation) compared to a country with no secessionist regions. Appendix Table C.8 shows the association between one of our model-based measures and the ten subcomponents of RAI. As can be seen, the statistically significant associations tend to be with components that are related to self-rule, rather than with shared-rule. This makes sense, given that secessionism is more concerned with self-determination than with participating in national policy.

[24] The index consists of the following subcomponents: C1: Security Apparatus, C2: Factionalized Elites, C3: Group Grievance, E1: Economic Decline, E2: Uneven Economic Development, E3: Human Flight and Brain Drain, P1: State Legitimacy, P2: Public Services, P3: Human Rights and Rule of Law, S1: Demographic Pressures, S2: Refugees and IDPs, X1: External Intervention.

[25] Data available at https://garymarks.web.unc.edu/data/regional-authority-2/. Accessed on September 16, 2021.

Table 2: Predicted Demand for Secession and State Fragility, Regional Autonomy, and Conflict

| Panel A: Institutional | Fragile State Index (2006-2021) | | | | | | Regional Authority Index (1950-2016) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Rank | | | Index | | | Total | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| # Secessionist Regions (IHS) | 15.232*** | | | 6.321*** | | | 4.258*** | | |
| | (4.128) | | | (1.856) | | | (1.479) | | |
| Share of Pop. Pro-Secession (IHS) | | 78.964** | | | 38.495*** | | | 60.515** | |
| | | (30.695) | | | (13.219) | | | (27.599) | |
| At least 1 Region Pro-Secession | | | 24.114*** | | | 9.749*** | | | 7.912*** |
| | | | (7.061) | | | (3.150) | | | (2.822) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.46 | 0.42 | 0.45 | 0.43 | 0.42 | 0.43 | 0.16 | 0.19 | 0.18 |
| Observations | 167 | 167 | 167 | 167 | 167 | 167 | 89 | 89 | 89 |

| Panel B: Conflict | Intensity of Conflict (1997-2020) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | # Deaths (IHS) | | | # Events (IHS) | | | # Years (IHS) | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| # Secessionist Regions (IHS) | 1.558*** | | | 1.087*** | | | 0.281** | | |
| | (0.338) | | | (0.366) | | | (0.112) | | |
| Share of Pop. Pro-Secession (IHS) | | 3.280 | | | -0.529 | | | 0.665 | |
| | | (3.375) | | | (3.355) | | | (6.082) | |
| At least 1 Region Pro-Secession | | | 2.391*** | | | 1.565** | | | 0.510** |
| | | | (0.593) | | | (0.617) | | | (0.199) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.45 | 0.38 | 0.44 | 0.24 | 0.20 | 0.23 | 0.64 | 0.75 | 0.64 |
| Observations | 173 | 173 | 173 | 173 | 173 | 173 | 173 | 173 | 173 |

Notes: Table reports results from regressions of measures of state fragility, regional autonomy, and conflict on model-generated measures of the demand for secession. Regressions are run at the level of countries, and include World Bank region fixed effects as additional controls. IHS denotes variables have been transformed using the inverse hyperbolic sine transformation. Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

**Conflict.** As a further validation, we explore the association between the model-based demand for secession and conflict within countries. Our main data on conflict come from the Armed Conflict Location and Event Data-ACLED (Raleigh et al. 2010). Panel B in Table 2 shows that most of our measures are positively and significantly associated with conflict intensity within countries. For example, countries that we predict have at least one secessionist region have on average 239% more deaths, are involved in 157% more conflict events, and suffer from conflict for 51% more years. Using additional conflict data from the UCDP/PRIO Armed Conflict Dataset (Gleditsch et al., 2002)) and terrorism data from the

Figure 9: Predicted Demand for Secession and Conflict Intensity (Deaths) across Countries

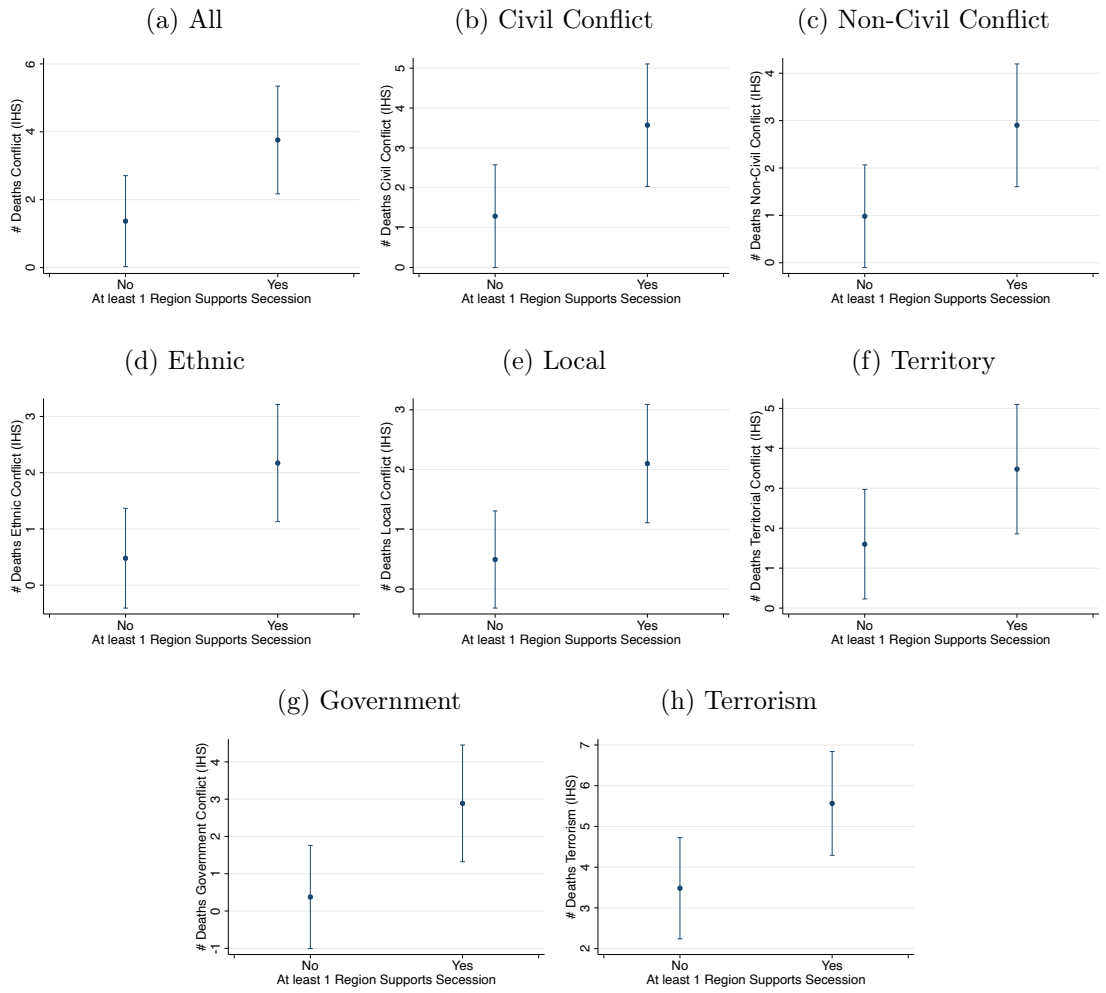



Figure shows the association between our model-based predicted demand for secession as measured by a country having at least 1 region in favor of secession and the number of deaths by type and source of conflict. IHS denotes variables have been transformed using the inverse hyperbolic sine transformation.

Global Terrorism Database (GTD),[26] Figure 9 shows that the results hold for various types of conflict, including civil, ethnic, and local, as well as for terrorism events.

**Alternative calibration.** Appendix Tables C.14-C.18 and Appendix Figure C.5 report the same validation analysis for our alternative calibration. The results are qualitatively similar: our predicted measures of secessionism are positively and significantly associated with actual secessionist activity, state fragility, regional autonomy, and conflict.

[26]GTD is available at https://www.start.umd.edu/gtd/. Analysis uses the September 2019 version accessed on November 6, 2019.

## 6.3 The Predicted Disintegration of the Soviet Union

As a final validation exercise, we analyze whether our calibrated model would have predicted the disintegration of the Soviet Union. Between 1990 and 1991, the 15 different republics that made up the Soviet Union became independent countries. Using data from the end of the 1980s, we compute for each of the republics the share of the population in favor of secession.

It is not immediately obvious how to get reliable data on income per capita from the Soviet era. In principle, we could use the G-Econ 4.0 data, because they go back to 1990. However, for most Soviet republics, the 1990 data are based on 1995 or later, and projected backward using a variety of methods. Given this drawback, we rely on the monograph by Flakierski (1992) who reports wage data of manual and non-manual labor for each Soviet republic in 1988. Since wages are not the same as income per capita, we use data on real GDP per capita from Russia in 1990 to rescale the wage data. As long as the ratio between GDP per capita and wages does not differ much across republics, this gives us a reasonable proxy of income per capita for 1988. For population data by republic, we use the 1989 Soviet Census.

Table 3: Predicted Secession of Soviet Republics

| Model-Predicted Share of Population in Favor of Secession from the Soviet Union | | | | | | | |
|---|---|---|---|---|---|---|---|
| Republic Wage Data 1988 | | Regional Wage Data 1988 | | GDP per Capita Data 1990 | | Alternative Calibration | |
| (1) Country | (2) Share | (3) Country | (4) Share | (5) Country | (6) Share | (7) Country | (8) Share |
| Uzbekistan | 0.02 | Uzbekistan | 0.02 | Armenia | 0.00 | Belarus | 0.00 |
| Belarus | 0.25 | Tajikistan | 0.03 | Azerbaijan | 0.00 | Kyrgyzstan | 0.01 |
| Kazakhstan | 0.46 | Belarus | 0.25 | Belarus | 0.00 | Ukraine | 0.02 |
| Kyrgyzstan | 0.54 | Kazakhstan | 0.46 | Kyrgyzstan | 0.00 | Kazakhstan | 0.43 |
| Latvia | 0.66 | Kyrgyzstan | 0.54 | Moldova | 0.00 | Latvia | 0.66 |
| Moldova | 0.73 | Moldova | 0.64 | Tajikistan | 0.00 | Moldova | 0.73 |
| Estonia | 0.75 | Latvia | 0.66 | Turkmenistan | 0.00 | Tajikistan | 0.75 |
| Tajikistan | 0.75 | Ukraine | 0.72 | Uzbekistan | 0.00 | Estonia | 0.75 |
| Ukraine | 0.77 | Estonia | 0.75 | Kazakhstan | 0.09 | Georgia | 0.78 |
| Georgia | 0.78 | Georgia | 0.78 | Ukraine | 0.66 | Uzbekistan | 0.81 |
| Azerbaijan | 0.86 | Azerbaijan | 0.86 | Latvia | 0.66 | Azerbaijan | 0.86 |
| Lithuania | 0.87 | Lithuania | 0.87 | Estonia | 0.75 | Lithuania | 0.87 |
| Turkmenistan | 0.91 | Turkmenistan | 0.91 | Georgia | 0.76 | Turkmenistan | 0.91 |
| Armenia | 0.95 | Armenia | 0.95 | Lithuania | 0.87 | Armenia | 0.95 |

Notes: Columns (1) and (2) are based on republic-level wage data of 1988 from Flakierski (1992). They have been scaled by the ratio of income per capita to wages in Russia in 1990. Columns (3) and (4) are based on the same data, but allow for within-republic wage differences between administrative regions by using data on relative income per capita of 1990 from G-Econ 4.0. Columns (5) and (6) are based on 1990 income per capita data from G-Econ 4.0. Columns (7) and (8) are based on the same data as columns (1) and (2), but use the parameters of the alternative calibration.

Using these data, what does the calibrated model predict? Columns (1) and (2) of Table 3 show a majority in favor of secession in all but three Soviet republics: Uzbekistan, Belarus, and Kazakhstan.[27] We conduct two further robustness checks. First, our income per capita proxy is at the level of each republic, because the Flakierski (1992) wage data are at the level of republics. To generate wage differences between subregions of the republics, we set the relative wage levels of the different subregions to those of income per capita of 1990 in G-Econ 4.0. The results are largely unchanged. Columns (3) and (4) show that one additional republic, Tajikistan, now prefers to stay within the Soviet Union. Second, we directly use the income per capita data from G-Econ 4.0. While these data have the drawback of being backward projected, they have the advantage of including all sources of income, and not just wages. Using these alternative income data, five more republics prefer to remain within the Soviet Union: Armenia, Azerbaijan, Kyrgyzstan, Moldova and Turkmenistan. In all three exercises, the three Baltic republics (Lithuania, Letonia, and Estonia), Georgia, and Ukraine prefer to leave the Soviet Union. And in all three exercises, Uzbekistan, Belarus, and Kazakhstan prefer to remain within the union.

How do these results compare to the historical record? One basic finding is that our model predicts that the Soviet Union was clearly unstable. Given that in our quantitative model only 5.9% of the 3,003 subnational regions have a majority in favor of secession, this was not a foregone conclusion.

To further investigate how well the model fits the historical record, we also compare our findings to the timeline of the breakup of the Soviet Union, and to the emergence of alliances in the post-Soviet world. In terms of timeline, in March 1990, Lithuania became the first republic to declare independence from the Soviet Union. The other two Baltic states, Latvia, and Estonia, followed suit in the subsequent months. The first non-Baltic republic to secede was Georgia in April 1991. In the ensuing Summer and Fall, most remaining republics became independent, with only Belarus, Russia, and Kazakhstan remaining in the union. Those last three finally also became independent countries in December 1991. Our findings are remarkably consistent with this timeline: the first four republics to declare independence have a majority in favor of secession in all of our exercises, and the two republics to remain in the union until the very end fail to reach a majority in all three exercises.

In the post-Soviet era, many cooperation agreements surfaced between subsets of the successor states. Perhaps the two most important ones are the Commonwealth of Independent States (CIS) and the Eurasian Economic Union (EAEU). The CIS is an intergovernmental organization that promotes economic, political, and military cooperation between its mem-

[27] We did not assess the propensity of Russia to secede. As the central player of the Soviet Union, its incentives to keep the union together were different from those of the other republics.

ber states. Its current members are Armenia, Azerbaijan, Belarus, Kazakhstan, Kyrgyzstan, Moldova, Russia, Tajikistan, and Uzbekistan. The EAEU is an economic union between Armenia, Belarus, Kazakhstan, Kyrgyzstan, and Russia. Consistent with what our findings would suggest in all three exercises, Georgia and Ukraine do not participate in either of these organizations, whereas Kazakhstan and Belarus are members of both.

As a last robustness check, we redo our first exercise using the parameters of the alternative calibration. In this case, the model predicts that all but four republics want to secede. As before, the first four republics to declare independence (Latvia, Lithuania, Estonia, and Georgia) have a majority in favor of independence, and the last two republics to remain in the union (Belarus and Kazakhstan) have less than 50% supporting secession. The only significant difference with our other exercises is that Ukraine now chooses to remain inside the Soviet Union. The reason for this difference is straightforward: under the alternative calibration, Ukrainian and Russian are aggregated into the same linguistic identity group, weakening the incentive to secede.

Overall, from the different comparisons with the data, and from our analysis of the Soviet Union, we can conclude that our quantitative model performs well in predicting potential secessionist activity around the globe.

# 7 Conclusions

This paper analyzed whether the demand for secession is mostly driven by economic or ethnolinguistic identity differences. To study this question, we proposed a model where the tax rate is determined by majority vote and where the utility an individual derives from the public good is declining in how distinct her identity is from the rest of the country. In equilibrium, subnational regions that are either richer or more distinct in terms of identity pay a higher tax rate than its residents would like.

Taking the model to the data for 3,003 subnational regions covering the globe, we generated measures of the instability of countries and the propensity to secede of subnational regions. To validate our model's predictions, we constructed a novel worldwide database of active secessionist movements at the subnational level, and found that our model-based measures of instability are strongly associated with observed secessionist activity. Lending further credibility to the model, we also showed a strong association of our measures with the fragility of states, violent conflict, and regional autonomy. In addition, the model is successful at predicting the breakup of the Soviet Union.

We used counterfactual analysis to gauge the relative importance of income per capita and linguistic identity in driving the demand for independence. We did so in three ways.

First, we compared what happens to the propensity to secede when removing either income per capita or linguistic identity differences, and found that the demand for independence is mostly driven by identity differences. Second, we evaluated by how much income per capita would have to drop for secessionist regions to give up on their demands. We found that large drops in income are needed, suggesting that economic differences only matter if they are large. Third, we assessed the effect of removing within-region linguistic identity differences, finding an important increase in the drive for independence. Overall, these different exercises strongly suggest that linguistic identity trumps income in determining secessionist tendencies.

The framework we have presented could be extended in several ways. First, we have focused on the propensity to secede by first-level administrative regions. A more complex analysis might also consider the possible secession of coalitions of first-level administrative regions or of subdivisions of those regions. Moreover, in some cases secession might be driven by the desire to join another country, rather than by becoming independent. Our framework can be used to study these questions as well.

Second, our model does not explicitly allow for fiscal decentralization. In that sense, we might want to interpret the demand for secession as more generally the demand for autonomy, including fiscal decentralization. Consistent with this interpretation, we have shown that the model-generated measures of secessionism are correlated with measures of regional autonomy. Future work should consider explicitly including the possibility of fiscal decentralization.

Third, the demand for secession may also depend on integration into higher-level institutions, such as the European Union. On the one hand, supranational organizations may make it easier to exit without losing the benefits of market size (Alesina, Spolaore and Wacziarg, 2000). On the other hand, supranational institutions may provide an alternative to exit, because they constrain nation states and limit their power (Gehring, 2021). Our framework could be extended and our data could be used to analyze these and other central questions about the stability of countries and the geopolitical organization of the world.

# References


[1] Akerlof, G. A. and E. Kranton, 2000. "Economics and Identity," *Quarterly Journal of Economics*, 15(3): 715-753.

[2] Alesina, A., Giuliano, P. and B. Reich, 2021. "Nation-Building and Education," *Economic Journal*, 131, 2273-2303.

[3] Alesina, A., Michalopoulos, S., and Papaioannou, E. (2016). "Ethnic inequality," *Journal of Political Economy*, 124(2), 428-488.

[4] Alesina, A. and E. Spolaore, 1997. "On the Number and Size of Nations," *Quarterly Journal of Economics*, 112(4): 1027-1056.

[5] Alesina, A. and E. Spolaore, 2003. *The Size of Nations*, Cambridge, MA: MIT Press.

[6] Alesina, A., Spolaore, E. and R. Wacziarg, 2000. "Economic Integration and Political Disintegration," *American Economic Review*, 90(5): 1276-1296.

[7] Álvarez Pereira, B., Portos, M. and J. Vourdas, 2018. "Waving Goodbye? The Determinants of Autonomism and Secessionism in Western Europe," *Regional Studies*, 52(2): 197-211.

[8] Blanc, G. and Kubo, M., 2021. "Schools, Language, and Nations: Evidence from a Natural Experiment in France," Brown University.

[9] Bolton, P. and G. Roland, 1997. "The Breakup of Nations: A Political Economy Analysis," *Quarterly Journal of Economics*, 112(4): 1057-1090.

[10] Comerford, D. and J. V. Rodríguez Mora, 2019. "The Gains from Economic Integration," *Economic Policy*, 34(98): 325-360.

[11] Desmet, K., Gomes, J. and I. Ortuño Ortín, 2020. "The Geography of Linguistic Diversity and the Provision of Public Goods," *Journal of Development Economics*.

[12] Desmet, K., Le Breton, M., Ortuño Ortín, I. and S. Weber, 2011. "The Stability and Breakup of Nations: A Quantitative Analysis," *Journal of Economic Growth*, 16, 183-213.

[13] Desmet, K., Ortuño Ortín, I. and R. Wacziarg, 2012. "The Political Economy of Linguistic Cleavages," *Journal of Development Economics*, 97, 322-338.

[14] Díez Medrano, J., 1994. "The Effects of Ethnic Segregation and Ethnic Competition on Political Mobilization in the Basque Country, 1988," *American Sociological Review*, 59(6): 873-889.

[15] Fearon, J. D., 2003. "Ethnic and Cultural Diversity by Country," *Journal of Economic Growth*, 8: 195-222.

[16] Flakierski, H., 1992. "Income Inequalities in the Former Soviet Union and Its Republics," *International Journal of Sociology*, 22(3): 1-87.

[17] Francois, P., Rainer, I. and F. Trebbi, 2015. "How is Power Shared in Africa?," *Econometrica*, 83(2): 465-503.

[18] Gancia, G., Ponzetto, G. and J. Ventura, 2022. "Globalization and Political Structure," *Journal of the European Economic Association*, 20(3): 1276-1310.

[19] Gehring, K., 2021. "Overcoming History Through Exit or Integration: Deep-Rooted Sources of Support for the European Union," *American Political Science Review*, 115(1): 199-217.

[20] Gehring, K. and S. Schneider, 2020. "Regional Resources and Democratic Secessionism," *Journal of Public Economics*, 181(C): S0047272719301343.

[21] Gershman, B. and D. Rivera, 2018. "Subnational Diversity in Sub-Saharan Africa: Insights from a New Dataset," *Journal of Development Economics*, 133: 231-263.

[22] Gleditsch, N. P., Wallensteen, P., Eriksson, M., Sollenberg, M. and H. Strand, 2002. 'Armed Conflict 1946-2001: A New Dataset," *Journal of Peace Research*, 39(5): 615-637.

[23] Gourevitch, P. A., 1979. "The Reemergence of 'Peripheral Nationalisms': Some Comparative Speculations on the Spatial Distribution of Political Leadership and Economic Growth," *Comparative Studies in Society and History*, 21(3): 303-322.

[24] Gurevich, T., Herman, P. R., Toubal, F. and Y. V. Yotov, 2020. "One Nation, One Language? Language Diversity, Trade and Welfare," unpublished manuscript.

[25] Gurr, T. R., 1993. *Minorities at Risk: A Global View of Ethnopolitical Conflicts*, Washington, DC: US Institute of Peace Press.

[26] Gurr, T. R., 2000. *Peoples versus States: Minorities at Risk in the New Century*, Washington, DC: US Institute of Peace Press.

[27] Hooghe, L., Marks, G., Schakel, A. H., Chapman-Osterkatz, S., Niedzwiecki, S. and S. Shair-Rosenfield, 2016. *Measuring Regional Authority. Volume I: A Postfunctionalist Theory of Governance,* Oxford: Oxford University Press.

[28] Hechter, M., 1975. *Internal Colonialism: The Celtic Fringe in British National Development, 1536-1966,* London, England: Routledge and Kegan Paul.

[29] Horowitz, D. L., 1981. "Patterns of Ethnic Separatism," *Comparative Studies in Society and History*, 23(2): 165-195.

[30] Huber, J. D. and L. Mayoral, 2019. "Group Inequality and the Severity of Civil Conflict," *Journal of Economic Growth*, 24(1): 1-41.

[31] Li, S. X., 2010. "Social Identities, Ethnic Diversity, and Tax Morale," *Public Finance Review*, 38(2): 146-177.

[32] Marshall, M. G. and T. R. Gurr, 2003. *Peace and Conflict 2003: A Global Survey of Armed Conflicts, Self-determination Movements, and Democracy*, College Park, MD: Center for International Development and Conflict Management.

[33] Marshall, M. G. and T. R. Gurr, 2005. *Peace and Conflict 2005: A Global Survey of Armed Conflicts, Self-determination Movements, and Democracy*, College Park, MD: Center for International Development and Conflict Management.

[34] Raleigh, C., Linke, A., Hegre, H. and J. Karlsen, 2010. "Introducing ACLED-Armed Conflict Location and Event Data," *Journal of Peace Research*, 47(5): 651-660.

[35] Rohner, D. and E. Zhuravskaya, 2023. *Nation Building: Big Lessons from Successes and Failures*, CEPR Press: London and Paris.

[36] Sambanis, N., Germann, M. and A. Schädel, 2018. "SDM: A New Data Set on Self-Determination Movements with an Application to the Reputational Theory of Conflict," *Journal of Conflict Resolution*, 62(3): 656-686.

[37] Sambanis, N. and B. Milanovic, 2014. "Explaining Regional Autonomy Differences in Decentralized Countries," *Comparative Political Studies*, 47(13): 1830-1855.

[38] Shayo, M., 2009. "A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution," *American Political Science Review*, 103(2): 147-174.

[39] Shayo, M., 2020. "Social Identity and Economic Policy," *Annual Review of Economics*, 12(1): 355-389.

[40] Sorens, J., 2005. "The Cross-Sectional Determinants of Secessionism in Advanced Democracies," *Comparative Political Studies*, 38: 304-326.

[41] Sorens, J., 2008. "Regionalists against Secession: The Political Economy of Territory in Advanced Democracies," *Nationalism and Ethnic Politics*, 14, 325-360.

[42] Vanschoonbeek, J., 2020. "Regional (In)stability in Europe: A Quantitative Model of State fragmentation," *Journal of Comparative Economics*, 48(3): 605-641.

[43] Weber, E., 1979. *Peasants into Frenchmen: Modernization of Rural France, 1870-1914,* Chatto and Windus: London.

# A Endogenous Identity Choice

In the model, the language an individual speaks is exogenously given. Hence, if we define an identity group as a collection of individuals defined by a common language, then an individual's identity is also exogenously given. Under this interpretation of the model, there is no endogenous identity choice. However, as we now show, we can reinterpret our model as one with endogenous identity, where an individual chooses to identify with either the nation or the subnational region. That is, once identity is defined by the nation or the subnational region, rather than by the language one speaks, our model becomes one with endogenous identity choice, in the spirit of the work by Shayo (2009, 2020).

Take an individual who speaks language $\ell$ and resides in region $r$ and country $C$. Recall that her utility is

$$v_C(\ell, r, C) = y(r, C)(1 - \tau(C)) + \alpha S(\ell, C)^{\delta}(\tau(C)Y(C))^{\beta} \tag{16}$$

where, using the terminology of Shayo (2020), we now refer to $S(\ell, C)^{\delta}$ as the individual's perceived proximity to the nation. Also recall that if region $r$ were to become an independent country, that same individual's utility would be

$$v_r(\ell, r, r) = y(r, r)(1 - \tau(r)) + \alpha S(\ell, r)^{\delta}(\tau(r)Y(r))^{\beta} \tag{17}$$

where $S(\ell, r)^{\delta}$ is the individual's perceived proximity to the region.

Now suppose that in addition to choosing a tax rate, an individual also chooses whether to identify with her country or with her subnational region. She makes this choice by comparing her current utility to the utility she would get if her subnational region were to become independent. Formally, an individual who speaks language $\ell$ and resides in region $r$ of country $C$ chooses the group $j$ that she identifies with by solving

$$\max_{j \in \{C,r\}} v_j(\ell, r, j) = \max_{j \in \{C,r\}} y(r, j)(1 - \tau(j)) + \alpha S(\ell, j)^{\delta}(\tau(j)Y(j))^{\beta}. \tag{18}$$

If an individual gains from her region $r$ becoming independent, she chooses to identify with subnational region $r$. If not, she identifies with country $C$. We model an individual's identity choice in the same way that we would model an individual's choice in a hypothetical election about independence. If in such a hypothetical election an individual would vote in favor of region $r$ becoming independent from country $C$, then we say that the individual identifies with subnational region $r$ rather than with country $C$.

The elements that affect identity choice are similar to those in Shayo (2009, 2020). First, the greater an individual's perceived proximity to group $j$, given by $S(\ell, j)$, the more likely she is to identify with group $j$. The greater the share of people in group $j$ that speak language $\ell$, the bigger the individual's perceived proximity to group $j$. As in Shayo (2009, 2020), this perceived proximity is exogenous. In our case, this exogeneity reflects the assumption that an individual's language is given. Second, the greater the material payoff or the status of group $j$, the more likely she is to identify with group $j$. Here, the material payoff is given by the amount of public goods provided by group $j$, $\tau(j)Y(j)$.

There are some differences with the standard models of endogenous identity. In Shayo (2009, 2020), when an individual chooses to identify with a specific group, she gets the actual utility associated with pertaining to that group. Here, instead, the actual utility an

individual gets is independent of her identity choice. If an individual chooses to identify with the subnational region, rather than with the country, she does so because of the higher utility she *would* get if her region became independent. However, as long as the region does not secede, she would continue to get the utility from residing in country $C$, rather than the utility associated with $r$ becoming independent.[28] As an example, suppose that a Catalan speaker in Catalonia would have a higher utility from being independent than from remaining in Spain. In that case, she would identify with Catalonia, rather than with Spain. However, the actual utility she gets would depend on the actual situation. As long as Catalonia continues to be part of Spain, she would get the utility associated with residing in Spain.

This alternative interpretation with endogenous identity choice changes how we interpret our counterfactual experiments. In the main body of the paper, we associate the language an individual speaks with the identity groups she belongs to. Hence, when asking whether linguistic differences or income differences matter more for secessionism, we refer to this question as assessing whether identity or income matters more for secessionism. Under the alternative interpretation where individuals identify with either the nation or the subnational region, rather than with the linguistic group, we refer to this same question as assessing whether language/ethnicity or income matters more for secessionism.

[28]This is not unlike how we might imagine the thought process that underlies an individual's choice to identify with a certain political party or with a certain ideology. An individual compares how her utility would differ depending on which party or which ideology would govern, and then identifies with the one that would give her the highest utility. The actual utility she gets would still be determined by whatever party or ideology actually governs, but her identity choice would be determined by comparing different potential utility levels.

# Online Appendix

# B Data Sources

**First-level administrative regions.** Global geographic distribution of administrative areas of all countries, at first levels of sub-division. Source: The Global Administrative Division Mapping project - GADM, version 3.6. available at https://gadm.org/index.html.

**Population.** Landscan at the 30″ by 30″ level for the year 2000, aggregated up to first-level administrative regions. Source: Bright and Coleman (2001).

**Income.** G-Econ 4.0 at the 1° by 1° level for the year 2000, aggregated up to first-level administrative regions. If a 1° by 1° grid cell overlaps more than one first-level administrative region, the income of that grid cell gets allocated to the different regions using their respective population shares, computed using Landscan population data at the 30″ by 30″ level. Source: Nordhaus et al. (2006).

**Language composition.** Language use data at the 5 km by 5 km level, aggregated up to first-level administrative regions. Source: Desmet, Gomes and Ortuño Ortín (2020).

**Secessionist movements.** Number, presence, and interest in all active secessionist movements around the world. Using Python, we scraped all the Wikipedia pages about active separatists movements in the world. For each movement we have information on its type (e.g., political party, militant, civil, or social organization,...), the link to any existing Wikipedia entries on it, and the name and location of the region it is associated with. This provides us with a list of 2,529 active secessionist movements. For each of these secessionist movements we then identify its geographical location at the first-level administrative region across countries. Specifically, we link each movement to all first-level administrative regions as defined in GADM v.36 for which it is trying to obtain autonomy, secession, or independence. Since the actual region associated with the movement may not always overlap perfectly with only one administrative region, we link it to all administrative regions that include or intersect the actual secessionist region. Whenever Wikipedia provides information on the administrative regions linked to the movement and these are the same as presented in GADM v.36, we directly use those. For all others, we use the location data of the proposed state, independent/autonomous region, or equivalent as provided by Wikipedia to identify the first-level administrative regions in GADM v.36 that intersect or contain it. Additionally, using Python, we scraped information on the number of visitors to all the Wikipedia entries of each movement in all languages over the 2015-2020 period. If a movement had no Wikipedia entries we assigned zero views to it. Source: List of active secessionist movements from Wikipedia at https://en.wikipedia.org/wiki/Lists_of_active_separatist_movements accessed on October 12, 2020. Wikipedia lists all movements active by continent (Africa, Asia, Europe, North America, Oceania, South America). Authors' computations.

**Fragile State Index (FSI).** Fragile States Index (FSI), calculated by the Fund for Peace. FSI is based on a conflict assessment framework that was developed by the Fund for Peace for assessing the vulnerability of states to collapse. Their framework was originally designed to measure this vulnerability and assess how it might affect projects in the field, and continues to be used widely by policy makers, field practitioners, and local community networks. The methodology uses both qualitative and quantitative indicators, relies on public source data, and produces quantifiable results. Twelve conflict risk indicators are used to measure the condition of a state at any given moment. The indicators provide a snapshot in time that can be measured against other snapshots in a time series to determine whether conditions are improving or worsening. Specifically, it is based on the following subcomponents: C1: Security Apparatus, C2: Factionalized Elites, C3: Group Grievance, E1: Economic Decline, E2: Uneven Economic Development, E3: Human Flight and Brain Drain, P1: State Legitimacy, P2: Public Services, P3: Human Rights and Rule of Law, S1: Demographic Pressures, S2: Refugees and IDPs, X1: External Intervention. Source: The data is accessible at https://fragilestatesindex.org/. Accessed on June 23, 2021.

**Regional Authority Index (FAI).** The Regional Authority Index (RAI) measures the authority in self rule and shared rule exercised by regional governments within their countries collected by Hooghe et al. (2016). It includes 96 countries (including China, India, Pakistan, Ukraine), has a temporal coverage from 1950 through 2018, and covers metropolitan and indigenous regions alongside conventional regions. Scoring is annual and the unit of analysis is the individual region. Regional authority is conceived as composed of self-rule (the authority exercised by a regional government over those who live in the region) and shared rule (the authority exercised by a region or its representatives in the country as a whole). Each domain is disaggregated in five dimensions that estimate fiscal, administrative, political, and constitutional authority. Source: Data available at https://garymarks.web.unc.edu/data/regional-authority-2/. Accessed on September 16, 2021.

**Conflict data.** Measures of the number of deaths from various types of conflict and terrorist attacks as provided by ACLED, UCDP/PRIO, and GTD at the first-level administrative regions across the world. Additionally, the number of years of conflict and number of conflict events for various conflict types at the first-level administrative regions across the world. We employ the types of conflict usually employed in the literature (Depetris-Chauvin and Özak, 2020; Moscona et al., 2020). Types of conflict based on ACLED data are: all, civil (where one of the actors must be the government), non-civil (where none of the actors can be the government), ethnic (where the conflict is driven by ethnic animosity, identified by all conflicts where the word "ethnic" appears in ACLED's actor names, notes, or associated actors), local (where only local actors are involved, this includes all conflicts where the ACLED interaction code is in $\{40, 41, \ldots, 48, 50, \ldots, 58, 60, \ldots, 67\}$). Using UCDP/PRIO we also construct measures for the following types of conflict: state-based (where one of the sides is the government), non-state based (where the government does not take part in conflict), territory (includes only conflicts over territorial disputes as identified by MILC), government (includes only conflicts over the type of political system, the replacement of the central government or the change of its composition, as identified by MILC). We use GTD to construct measures due to terrorist

attacks. Sources: UCDP/PRIO Georeferenced Event Dataset (GED) v20.1(Pettersson et al. 2021). Data available at https://ucdp.uu.se/downloads/. Accessed on September 19, 2020. UCDP Managing Intrastate Low-intensity Conflict (MILC) dataset v.10 (Melander et al. 2009). Data available at https://ucdp.uu.se/downloads/. Accessed on September 19, 2020. Armed Conflict Location and Event Data-ACLED (Raleigh et al. 2010). Data available at https://acleddata.com/data-export-tool/. Accessed on September 17, 2020. Global Terrorism Database (GTD). GTD is available at https://www.start.umd.edu/gtd/. Analysis uses the September 2019 version accessed on November 6, 2019. Authors' computations.

**Additional data references.**

Bright, E. and P. R. Coleman, 2001. *Landscan 2000,* Oak Ridge, TN: Oak Ridge National Laboratory.

Depetris-Chauvin, E., and Ö. Özak, 2020. "Borderline Disorder: (De facto) Historical Ethnic Borders and Contemporary Conflict in Africa," unpublished manuscript.

Nordhaus, W., Azam, Q., Novoa Corderi, D., Hood, K., Victor, N., Mohammed, M., Miltner, A. and J. Weiss, 2006. "The G-Econ Database on Gridded Output: Methods and Data," New Haven, CT: Yale University.

Melander, E., Möller, F. and M. Öberg, 2009. "Managing Intrastate Low-Intensity Armed Conflict 1993-2004: A New Dataset," *International Interactions*, 35(1): 58-85.

Moscona, J., Nunn, N. and J. A. Robinson, 2020. "Segmentary Lineage Organization and Conflict in Sub-Saharan Africa," *Econometrica*, 88(5): 1999-2036.

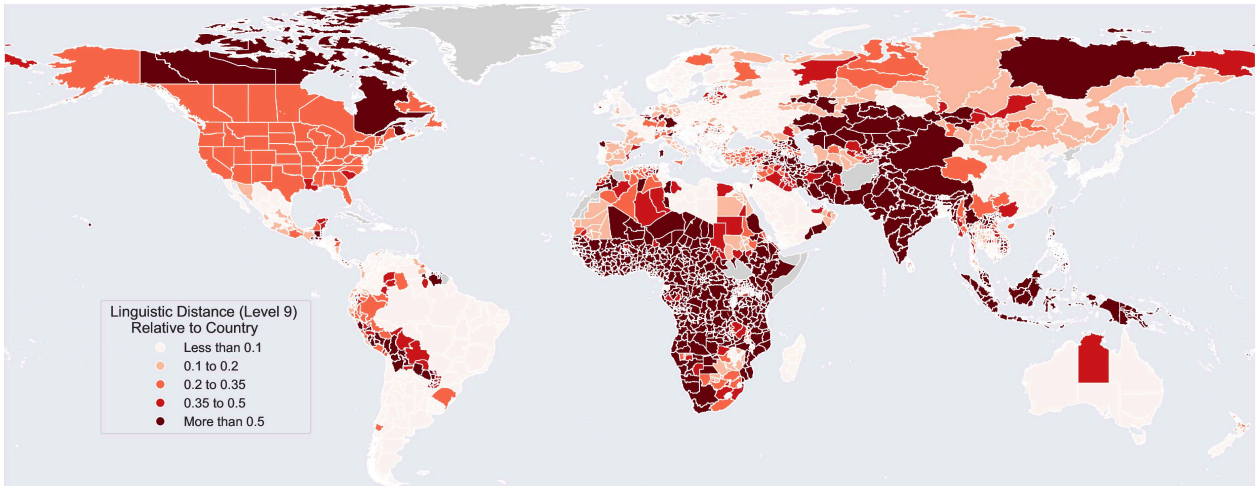
Pettersson, T., Davis, S., Deniz, A., Engström, G., Hawach, N., Högbladh, S., Sollenberg, M. and M. Öberg, 2021. "Organized Violence 1989-2020, with a Special Emphasis on Syria," *Journal of Peace Research*, 58(4): 809-825.

# C Additional Tables and Figures

## C.1 Linguistic Distances at Different Levels of Aggregation

Figure C.1: Linguistic Distance between Subnational Regions and Countries: Levels of Aggregation 9 and 2

(a) Aggregation Level 9




(b) Aggregation Level 2

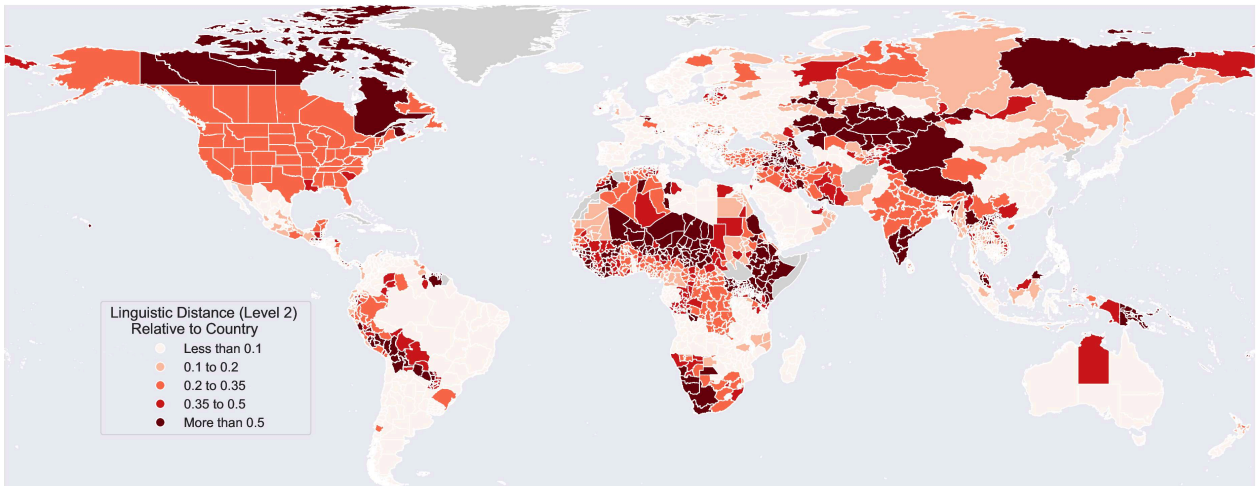
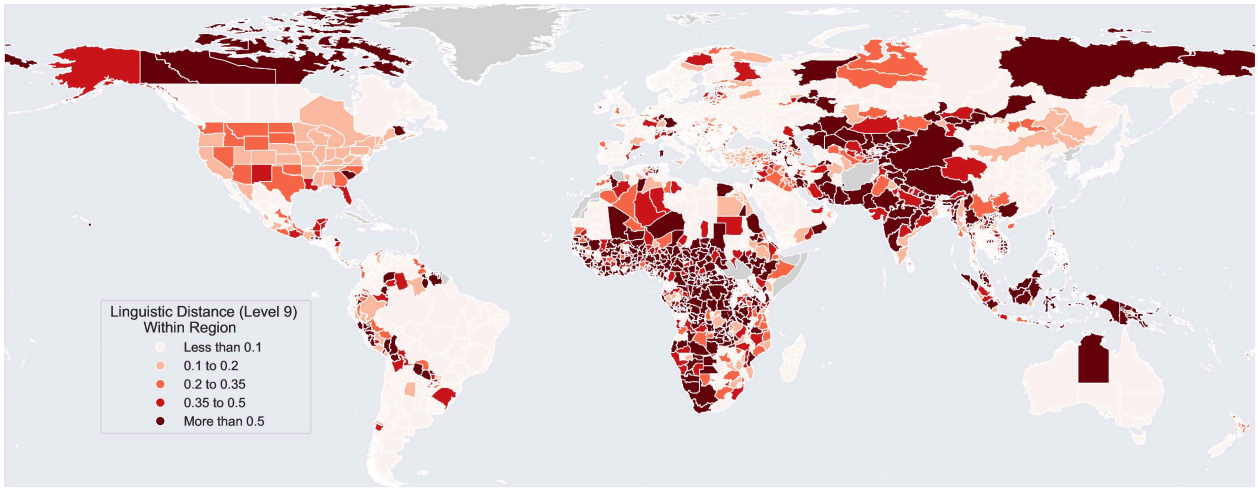



Figure depicts the linguistic distance between each subnational region and the country it belongs to. It is measured as the expected distance between a randomly drawn individual of the subnational region and a randomly drawn individual of the country. Darker colors indicate greater linguistic distances between region and country, and hence a stronger propensity to secede. Languages are defined at either intermediate levels of coarseness (aggregation level 9) or at high levels of coarseness (aggregation level 2). For example, at aggregation level 2, Catalan and Spanish belong to the same group, but Dutch and French do not.

Figure C.2: Within-Subnational Region Linguistic Distance: Levels of Aggregation 9 and 2

(a) Aggregation Level 9




(b) Aggregation Level 2

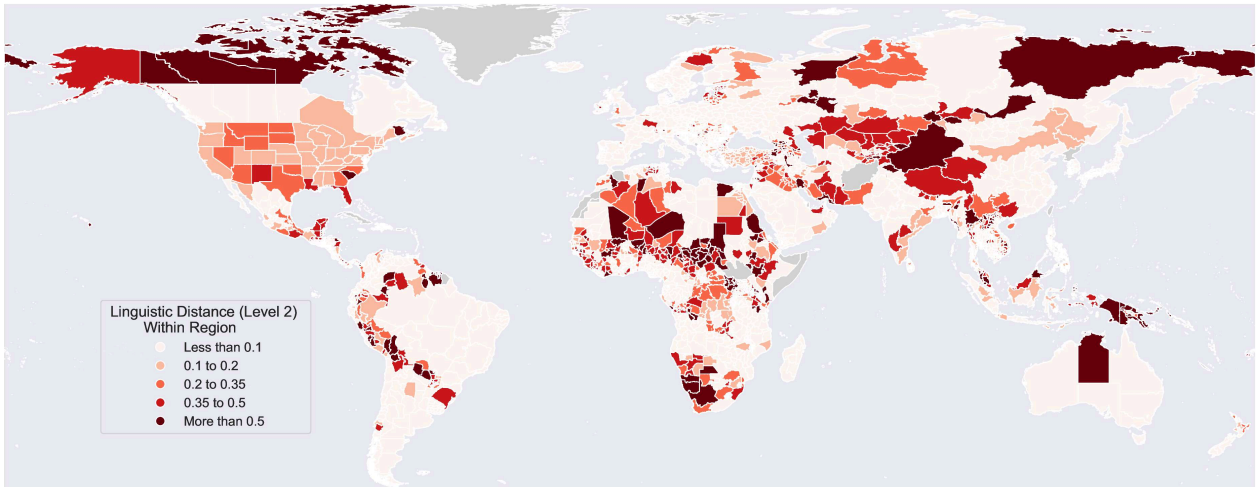



Figure depicts the linguistic distance between each subnational region and itself. It is measured as the expected distance between two randomly drawn individuals of the region. Darker colors indicate a greater within-region distance, and hence a weaker propensity to secede. Languages are defined at either intermediate levels of coarseness (aggregation level 9) or at high levels of coarseness (aggregation level 2). For example, at aggregation level 2, Catalan and Spanish belong to the same group, but Dutch and French do not.

## C.2 Language and Religion

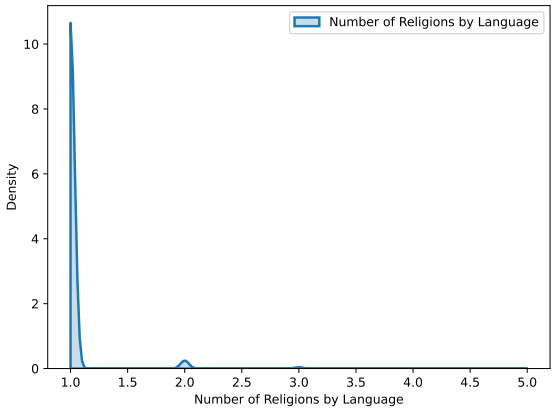
While we focus on language as the identity trait that defines a group, in principle we could include additional identity traits, such as religion. In the main text, we already mentioned that there are only a handful of secessionist regions with a distinct religion that do not also have a distinct language: Zanzibar, Northern Ireland, and Bangsamoro. In other regions where religion may be a motive of secession, language is also different. Examples include Tibet, Northern Cyprus, Nagorno-Karabakh, and Punjab. In that sense, our focus on linguistic identity is reasonable because it defines these regions as having a distinct identity.

To further explore the role of language and religion for group identity, we use data from the Joshua Project (https://joshuaproject.net/), a research initiative seeking to identify "the ethnic people groups of the world with the fewest followers of Christ". The project identifies all peoples in all countries and provides information on their primary language and religion. For each country, we compute the number of languages spoken by each religious denomination, and the number of religions followed by the speakers of each language. Figures C.3a and C.3b show the distributions of the number of languages spoken by people who practice the same religion, and the number of religions practiced by speakers of a language within the country. As we can see, most coreligionist within a country tend to speak different languages (up to 92). In contrast, people who speak the same language in a country also tend to practice the same religion. That is, religious groups are linguistically heterogeneous, whereas language groups are religiously homogeneous. This suggests that language groups are more cohesive than religious groups, making it more likely for language to define group identity.

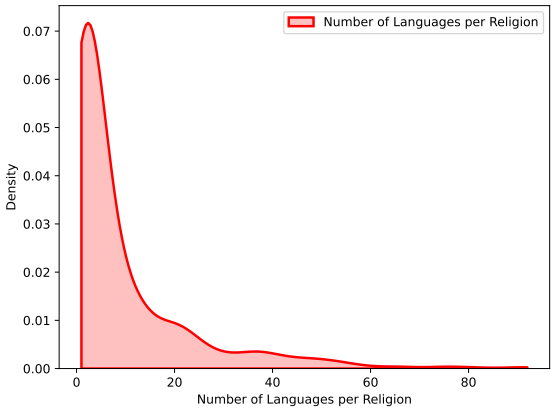
One potential concern is that these results include recent migrants, such as Americans living in Turkey. To improve the analysis, we restrict ourselves to people groups that are native to the country. For example, in Turkey there are three primary religions practiced by indigenous peoples: Christianity, Islam, and Ethnic Religions. Followers of Christianity belong to two language groups (Hertevin, Turoyo), of Islam to 13 language groups (Abkhaz, Northern and Southern Zazaki, Domari, Kabardian, Northern Kurdish, Laz, Persian, Iranian, Bulgarian, Pontic, Balkan Gagauz Turkish, Turkish, Levantine Arabic), and of Ethnic Religions to two language groups (Turkish, Northern Kurdish). In contrast, there are 16 language groups, all of which follow just one religion, with the exception of speakers of Turkish and Kurdish, which have two primary religions (Islam and Ethnic Religions). Figures C.3c and C.3d show the distributions of the number of languages spoken by indigenous people who practice the same religion, and the number of religions practiced by speakers of a language within the country. The results again show that most coreligionists within a country tend to speak different languages (up to 86), while members of a language group tend to practice the same religion (96% of language groups within countries practice a unique religion). As a result, religious groups are heterogeneous, whereas language groups are homogeneous, suggesting that language is more likely to provide group cohesiveness.

Figure C.3: Within-Country Association of Number of Languages and Religions

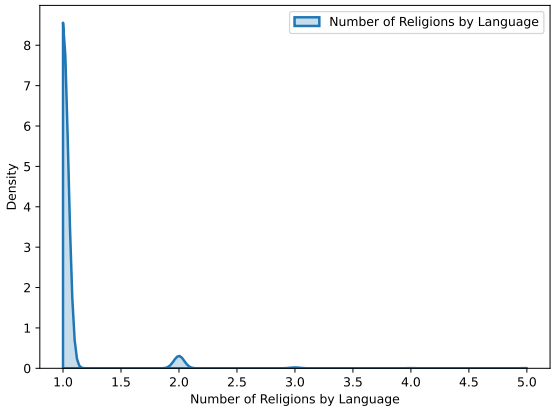
(a) Number of Religions by Language (within countries)




(b) Number of Languages by Religion (within countries)




(c) Number of Religions by Language (within countries, Indigenous Peoples)




(d) Number of Languages by Religion (within countries, Indigenous Peoples)

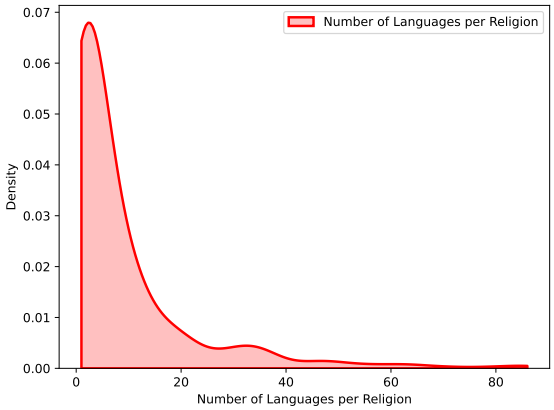



Figure plots the distribution of the number of religions by language groups within countries (Panels (a) and (c)) and the distribution of the number of languages by religion within countries (Panels (b) and (d)). Panels (a) and (b) take into account all people groups (including recent immigrants) whereas Panels (c) and (d) only take into account people groups that are native to a country. The plots show that, within countries, language groups tend to be religiously homogeneous, whereas religious groups tend to be linguistically heterogeneous.

## C.3 Propensity to Secede by Country: Baseline Calibration

Table C.1: Propensity to Secede by Country: Baseline Calibration

| | | Population | | Share | | | Population | | Share | | | Population | | Share |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Nat | Reg | Reg | | | Nat | Reg | Reg | | | Nat | Reg | Reg |
| 1 | Micronesia | 0.678 | 0.702 | 0.500 | 59 | Belize | 0.059 | 0.096 | 0.167 | 117 | Denmark | 0.005 | 0.000 | 0.000 |
| 2 | South Sudan | 0.632 | 0.561 | 0.500 | 60 | Uganda | 0.055 | 0.056 | 0.069 | 118 | Bolivia | 0.004 | 0.000 | 0.000 |
| 3 | Comoros | 0.415 | 0.424 | 0.667 | 61 | Serbia | 0.053 | 0.000 | 0.000 | 119 | Kuwait | 0.004 | 0.000 | 0.000 |
| 4 | Vanuatu | 0.367 | 0.265 | 0.333 | 62 | Panama | 0.052 | 0.011 | 0.077 | 120 | Sweden | 0.004 | 0.000 | 0.000 |
| 5 | Papua NG | 0.363 | 0.348 | 0.409 | 63 | Philippines | 0.052 | 0.040 | 0.063 | 121 | Tunisia | 0.004 | 0.000 | 0.000 |
| 6 | India | 0.336 | 0.342 | 0.314 | 64 | Yemen | 0.049 | 0.034 | 0.095 | 122 | Mauritania | 0.004 | 0.000 | 0.000 |
| 7 | Cameroon | 0.335 | 0.203 | 0.200 | 65 | Russia | 0.047 | 0.042 | 0.036 | 123 | Macedonia | 0.004 | 0.000 | 0.000 |
| 8 | Togo | 0.319 | 0.129 | 0.200 | 66 | Finland | 0.045 | 0.000 | 0.000 | 124 | USA | 0.004 | 0.004 | 0.020 |
| 9 | Tanzania | 0.319 | 0.310 | 0.333 | 67 | Madagascar | 0.045 | 0.000 | 0.000 | 125 | Nicaragua | 0.003 | 0.000 | 0.000 |
| 10 | Benin | 0.274 | 0.280 | 0.333 | 68 | Syria | 0.043 | 0.000 | 0.000 | 126 | Slovakia | 0.003 | 0.000 | 0.000 |
| 11 | Solomon Isl | 0.254 | 0.355 | 0.500 | 69 | Kosovo | 0.043 | 0.000 | 0.000 | 127 | Norway | 0.003 | 0.000 | 0.000 |
| 12 | Côte d'Ivoire | 0.235 | 0.195 | 0.214 | 70 | Senegal | 0.043 | 0.046 | 0.071 | 128 | Poland | 0.002 | 0.000 | 0.000 |
| 13 | Chad | 0.229 | 0.199 | 0.174 | 71 | Malawi | 0.043 | 0.031 | 0.074 | 129 | Australia | 0.002 | 0.000 | 0.000 |
| 14 | Sudan | 0.224 | 0.230 | 0.167 | 72 | Uzbekistan | 0.042 | 0.000 | 0.000 | 130 | Sri Lanka | 0.002 | 0.000 | 0.000 |
| 15 | Namibia | 0.202 | 0.184 | 0.154 | 73 | Laos | 0.042 | 0.000 | 0.000 | 131 | Slovenia | 0.002 | 0.000 | 0.000 |
| 16 | Pakistan | 0.198 | 0.230 | 0.250 | 74 | Libya | 0.041 | 0.040 | 0.091 | 132 | Latvia | 0.002 | 0.000 | 0.000 |
| 17 | Nigeria | 0.195 | 0.130 | 0.162 | 75 | France | 0.037 | 0.000 | 0.000 | 133 | Chile | 0.002 | 0.000 | 0.000 |
| 18 | DRC | 0.188 | 0.065 | 0.038 | 76 | Vietnam | 0.035 | 0.011 | 0.016 | 134 | El Salvador | 0.002 | 0.000 | 0.000 |
| 19 | Oman | 0.175 | 0.096 | 0.182 | 77 | Bhutan | 0.035 | 0.000 | 0.000 | 135 | Bulgaria | 0.002 | 0.000 | 0.000 |
| 20 | Indonesia | 0.171 | 0.132 | 0.303 | 78 | Peru | 0.034 | 0.000 | 0.000 | 136 | Iraq | 0.002 | 0.000 | 0.000 |
| 21 | CAR | 0.166 | 0.144 | 0.059 | 79 | Azerbaijan | 0.031 | 0.000 | 0.000 | 137 | Trinidad | 0.001 | 0.000 | 0.000 |
| 22 | Rep Congo | 0.159 | 0.021 | 0.083 | 80 | Suriname | 0.027 | 0.000 | 0.000 | 138 | Albania | 0.001 | 0.000 | 0.000 |
| 23 | Thailand | 0.158 | 0.152 | 0.078 | 81 | Mexico | 0.027 | 0.000 | 0.000 | 139 | Bosnia | 0.001 | 0.000 | 0.000 |
| 24 | Burkina Faso | 0.155 | 0.143 | 0.154 | 82 | Ecuador | 0.024 | 0.000 | 0.000 | 140 | Israel | 0.001 | 0.000 | 0.000 |
| 25 | Nepal | 0.154 | 0.000 | 0.000 | 83 | Guyana | 0.023 | 0.000 | 0.000 | 141 | Dom Rep | 0.001 | 0.000 | 0.000 |
| 26 | South Africa | 0.148 | 0.167 | 0.222 | 84 | Tajikistan | 0.021 | 0.031 | 0.200 | 142 | Ireland | 0.000 | 0.000 | 0.000 |
| 27 | Djibouti | 0.144 | 0.147 | 0.200 | 85 | Gambia | 0.020 | 0.000 | 0.000 | 143 | Argentina | 0.000 | 0.000 | 0.000 |
| 28 | China | 0.143 | 0.158 | 0.161 | 86 | Gabon | 0.020 | 0.000 | 0.000 | 144 | Cyprus | 0.000 | 0.000 | 0.000 |
| 29 | Zimbabwe | 0.140 | 0.058 | 0.100 | 87 | Egypt | 0.019 | 0.008 | 0.111 | 145 | UAE | 0.000 | 0.000 | 0.000 |
| 30 | Guinea | 0.140 | 0.190 | 0.125 | 88 | Brazil | 0.017 | 0.000 | 0.000 | 146 | Haiti | 0.000 | 0.000 | 0.000 |
| 31 | Eq Guinea | 0.138 | 0.003 | 0.143 | 89 | Jordan | 0.016 | 0.000 | 0.000 | 147 | Antigua | 0.000 | 0.000 | 0.000 |
| 32 | Mali | 0.135 | 0.000 | 0.000 | 90 | Lebanon | 0.015 | 0.000 | 0.000 | 148 | Bahrain | 0.000 | 0.000 | 0.000 |
| 33 | Ghana | 0.133 | 0.000 | 0.000 | 91 | Algeria | 0.014 | 0.001 | 0.021 | 149 | Belarus | 0.000 | 0.000 | 0.000 |
| 34 | Saudi Arabia | 0.131 | 0.135 | 0.077 | 92 | Honduras | 0.014 | 0.001 | 0.056 | 150 | Burundi | 0.000 | 0.000 | 0.000 |
| 35 | Bangladesh | 0.126 | 0.179 | 0.286 | 93 | Costa Rica | 0.014 | 0.000 | 0.000 | 151 | Cape Verde | 0.000 | 0.000 | 0.000 |
| 36 | Italy | 0.126 | 0.027 | 0.050 | 94 | Colombia | 0.014 | 0.005 | 0.031 | 152 | Estonia | 0.000 | 0.000 | 0.000 |
| 37 | Eritrea | 0.122 | 0.197 | 0.167 | 95 | Croatia | 0.014 | 0.000 | 0.000 | 153 | Grenada | 0.000 | 0.000 | 0.000 |
| 38 | Myanmar | 0.120 | 0.095 | 0.267 | 96 | Mongolia | 0.013 | 0.000 | 0.000 | 154 | Iceland | 0.000 | 0.000 | 0.000 |
| 39 | Guinea-Bissau | 0.116 | 0.148 | 0.222 | 97 | Cambodia | 0.012 | 0.011 | 0.080 | 155 | Jamaica | 0.000 | 0.000 | 0.000 |
| 40 | Iran | 0.115 | 0.143 | 0.129 | 98 | Hungary | 0.011 | 0.000 | 0.000 | 156 | Kazakhstan | 0.000 | 0.000 | 0.000 |
| 41 | Ethiopia | 0.114 | 0.033 | 0.273 | 99 | Romania | 0.011 | 0.000 | 0.000 | 157 | Lithuania | 0.000 | 0.000 | 0.000 |
| 42 | Germany | 0.112 | 0.013 | 0.063 | 100 | UK | 0.010 | 0.000 | 0.000 | 158 | Luxembourg | 0.000 | 0.000 | 0.000 |
| 43 | Kenya | 0.109 | 0.132 | 0.128 | 101 | Venezuela | 0.010 | 0.000 | 0.000 | 159 | Malta | 0.000 | 0.000 | 0.000 |
| 44 | Netherlands | 0.109 | 0.164 | 0.286 | 102 | Turkey | 0.010 | 0.000 | 0.000 | 160 | Mauritius | 0.000 | 0.000 | 0.000 |
| 45 | Spain | 0.108 | 0.061 | 0.056 | 103 | Czech Rep | 0.010 | 0.000 | 0.000 | 161 | Montenegro | 0.000 | 0.000 | 0.000 |
| 46 | Sierra Leone | 0.097 | 0.104 | 0.250 | 104 | Ukraine | 0.009 | 0.000 | 0.000 | 162 | Morocco | 0.000 | 0.000 | 0.000 |
| 47 | Belgium | 0.085 | 0.000 | 0.000 | 105 | Moldova | 0.009 | 0.000 | 0.000 | 163 | New Zealand | 0.000 | 0.000 | 0.000 |
| 48 | Mozambique | 0.083 | 0.025 | 0.091 | 106 | Kyrgyzstan | 0.009 | 0.000 | 0.000 | 164 | Rwanda | 0.000 | 0.000 | 0.000 |
| 49 | Georgia | 0.079 | 0.000 | 0.000 | 107 | Austria | 0.009 | 0.000 | 0.000 | 165 | St Kitts | 0.000 | 0.000 | 0.000 |
| 50 | Brunei | 0.073 | 0.000 | 0.000 | 108 | Greece | 0.009 | 0.000 | 0.000 | 166 | St Lucia | 0.000 | 0.000 | 0.000 |
| 51 | Angola | 0.073 | 0.016 | 0.056 | 109 | Paraguay | 0.008 | 0.009 | 0.056 | 167 | St Vincent | 0.000 | 0.000 | 0.000 |
| 52 | Fiji | 0.073 | 0.000 | 0.000 | 110 | Japan | 0.007 | 0.008 | 0.021 | 168 | Samoa | 0.000 | 0.000 | 0.000 |
| 53 | Malaysia | 0.070 | 0.000 | 0.000 | 111 | Turkmenistan | 0.007 | 0.000 | 0.000 | 169 | Seychelles | 0.000 | 0.000 | 0.000 |
| 54 | Liberia | 0.069 | 0.066 | 0.133 | 112 | Switzerland | 0.007 | 0.000 | 0.000 | 170 | South Korea | 0.000 | 0.000 | 0.000 |
| 55 | Zambia | 0.067 | 0.000 | 0.000 | 113 | Lesotho | 0.007 | 0.000 | 0.000 | 171 | Eswatini | 0.000 | 0.000 | 0.000 |
| 56 | Botswana | 0.066 | 0.007 | 0.067 | 114 | Portugal | 0.006 | 0.000 | 0.000 | 172 | Tonga | 0.000 | 0.000 | 0.000 |
| 57 | Guatemala | 0.065 | 0.000 | 0.000 | 115 | Armenia | 0.006 | 0.000 | 0.000 | 173 | Uruguay | 0.000 | 0.000 | 0.000 |
| 58 | Niger | 0.059 | 0.029 | 0.125 | 116 | Canada | 0.006 | 0.001 | 0.077 | | | | | |

This table provides for each country three measures of the propensity to secede: the share of the country's population in favor of secession, the share of the country's population living in regions with a majority in favor of secession, and the share of the country's regions with a majority in favor of secession. These predicted measures are based on the baseline calibration. Abbreviations used: Antigua: Antigua and Barbuda; Bosnia: Bosnia and Herzegovina; CAR: Central African Republic; Czech Rep: Czech Republic; DRC: Democratic Republic of the Congo; Dom Rep: Dominican Republic; Eq Ginea: Equatorial Guinea; Papua NG: Papua New Guinea; Rep Congo: Republic of Congo; Solomon Isl: Solomon Islands; St Kitts: Saint Kitts and Nevis; St Vincent: Saint Vincent and the Grenadines; Trinidad: Trinidad and Tobago; UAE: United Arab Emirates.

## C.4 Propensity to Secede by Country: Alternative Calibration

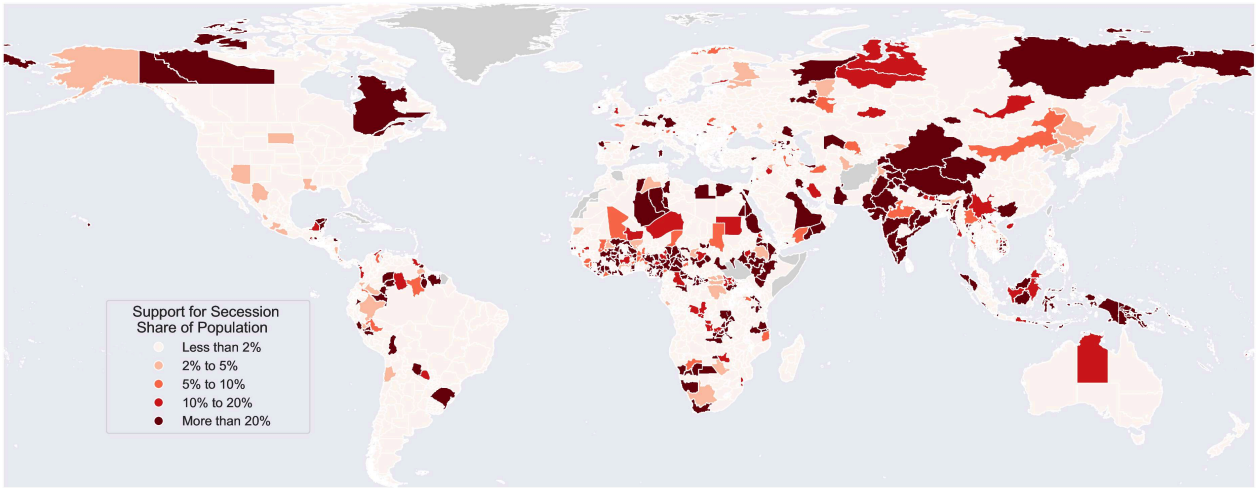
Table C.2: Propensity to Secede by Country: Alternative Calibration

| | | Population | | Share | | | Population | | Share | | | Population | | Share |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Nat | Reg | Reg | | | Nat | Reg | Reg | | | Nat | Reg | Reg |
| 1 | Pakistan | 0.674 | 0.767 | 0.375 | 59 | Kenya | 0.037 | 0.036 | 0.043 | 117 | USA | 0.003 | 0.004 | 0.020 |
| 2 | Papua NG | 0.569 | 0.622 | 0.636 | 60 | France | 0.036 | 0.000 | 0.000 | 118 | Macedonia | 0.003 | 0.000 | 0.000 |
| 3 | Belgium | 0.560 | 0.590 | 0.333 | 61 | Yemen | 0.035 | 0.034 | 0.095 | 119 | Greece | 0.003 | 0.000 | 0.000 |
| 4 | India | 0.458 | 0.496 | 0.400 | 62 | China | 0.033 | 0.017 | 0.065 | 120 | Norway | 0.003 | 0.000 | 0.000 |
| 5 | Ethiopia | 0.456 | 0.640 | 0.545 | 63 | Azerbaijan | 0.032 | 0.000 | 0.000 | 121 | Comoros | 0.002 | 0.000 | 0.000 |
| 6 | South Sudan | 0.381 | 0.373 | 0.500 | 64 | Mozambique | 0.030 | 0.000 | 0.000 | 122 | Nicaragua | 0.002 | 0.000 | 0.000 |
| 7 | Togo | 0.362 | 0.330 | 0.200 | 65 | Peru | 0.029 | 0.000 | 0.000 | 123 | Sri Lanka | 0.002 | 0.000 | 0.000 |
| 8 | Vanuatu | 0.301 | 0.240 | 0.333 | 66 | Suriname | 0.027 | 0.000 | 0.000 | 124 | Australia | 0.002 | 0.000 | 0.000 |
| 9 | Liberia | 0.298 | 0.325 | 0.333 | 67 | Zambia | 0.026 | 0.000 | 0.000 | 125 | Gabon | 0.002 | 0.000 | 0.000 |
| 10 | Burkina Faso | 0.268 | 0.196 | 0.231 | 68 | Guinea-Bissau | 0.026 | 0.011 | 0.111 | 126 | Latvia | 0.002 | 0.000 | 0.000 |
| 11 | Côte d'Ivoire | 0.260 | 0.241 | 0.214 | 69 | Vietnam | 0.025 | 0.011 | 0.016 | 127 | El Salvador | 0.002 | 0.000 | 0.000 |
| 12 | Chad | 0.239 | 0.199 | 0.130 | 70 | Serbia | 0.025 | 0.000 | 0.000 | 128 | Chile | 0.002 | 0.000 | 0.000 |
| 13 | Canada | 0.233 | 0.236 | 0.154 | 71 | Eq Guinea | 0.024 | 0.003 | 0.143 | 129 | Slovenia | 0.002 | 0.000 | 0.000 |
| 14 | CAR | 0.217 | 0.305 | 0.235 | 72 | Botswana | 0.021 | 0.007 | 0.067 | 130 | Trinidad | 0.001 | 0.000 | 0.000 |
| 15 | Philippines | 0.215 | 0.196 | 0.100 | 73 | Ireland | 0.021 | 0.000 | 0.000 | 131 | Bosnia | 0.001 | 0.000 | 0.000 |
| 16 | Namibia | 0.210 | 0.184 | 0.154 | 74 | Tajikistan | 0.021 | 0.031 | 0.200 | 132 | Albania | 0.001 | 0.000 | 0.000 |
| 17 | Nigeria | 0.193 | 0.188 | 0.216 | 75 | Italy | 0.021 | 0.027 | 0.050 | 133 | Israel | 0.001 | 0.000 | 0.000 |
| 18 | Indonesia | 0.174 | 0.140 | 0.212 | 76 | Guyana | 0.020 | 0.015 | 0.100 | 134 | Dom Rep | 0.001 | 0.000 | 0.000 |
| 19 | Mali | 0.174 | 0.000 | 0.000 | 77 | Brazil | 0.017 | 0.000 | 0.000 | 135 | Bulgaria | 0.001 | 0.000 | 0.000 |
| 20 | Thailand | 0.158 | 0.152 | 0.078 | 78 | Syria | 0.015 | 0.000 | 0.000 | 136 | Sweden | 0.001 | 0.000 | 0.000 |
| 21 | Benin | 0.150 | 0.184 | 0.250 | 79 | Lebanon | 0.015 | 0.000 | 0.000 | 137 | Slovakia | 0.001 | 0.000 | 0.000 |
| 22 | Djibouti | 0.144 | 0.147 | 0.200 | 80 | Honduras | 0.014 | 0.001 | 0.056 | 138 | Poland | 0.000 | 0.000 | 0.000 |
| 23 | Georgia | 0.144 | 0.098 | 0.083 | 81 | Rep Congo | 0.013 | 0.021 | 0.083 | 139 | Cyprus | 0.000 | 0.000 | 0.000 |
| 24 | Bolivia | 0.142 | 0.000 | 0.000 | 82 | Mexico | 0.013 | 0.000 | 0.000 | 140 | Argentina | 0.000 | 0.000 | 0.000 |
| 25 | Nepal | 0.138 | 0.000 | 0.000 | 83 | Mongolia | 0.013 | 0.000 | 0.000 | 141 | Angola | 0.000 | 0.000 | 0.000 |
| 26 | Sierra Leone | 0.136 | 0.104 | 0.250 | 84 | Ukraine | 0.012 | 0.000 | 0.000 | 142 | Antigua | 0.000 | 0.000 | 0.000 |
| 27 | Cameroon | 0.136 | 0.000 | 0.000 | 85 | Turkey | 0.012 | 0.017 | 0.012 | 143 | Bahrain | 0.000 | 0.000 | 0.000 |
| 28 | Saudi Arabia | 0.135 | 0.135 | 0.077 | 86 | Costa Rica | 0.012 | 0.000 | 0.000 | 144 | Belarus | 0.000 | 0.000 | 0.000 |
| 29 | Solomon Isl | 0.131 | 0.192 | 0.300 | 87 | Kazakhstan | 0.011 | 0.000 | 0.000 | 145 | Burundi | 0.000 | 0.000 | 0.000 |
| 30 | Oman | 0.122 | 0.000 | 0.000 | 88 | Niger | 0.011 | 0.000 | 0.000 | 146 | Cape Verde | 0.000 | 0.000 | 0.000 |
| 31 | Eritrea | 0.122 | 0.197 | 0.167 | 89 | Croatia | 0.011 | 0.000 | 0.000 | 147 | Estonia | 0.000 | 0.000 | 0.000 |
| 32 | Guinea | 0.113 | 0.190 | 0.125 | 90 | Egypt | 0.010 | 0.004 | 0.074 | 148 | Grenada | 0.000 | 0.000 | 0.000 |
| 33 | Germany | 0.112 | 0.013 | 0.063 | 91 | Romania | 0.010 | 0.000 | 0.000 | 149 | Haiti | 0.000 | 0.000 | 0.000 |
| 34 | Spain | 0.108 | 0.061 | 0.056 | 92 | Venezuela | 0.010 | 0.000 | 0.000 | 150 | Iceland | 0.000 | 0.000 | 0.000 |
| 35 | DRC | 0.101 | 0.085 | 0.077 | 93 | Cambodia | 0.010 | 0.011 | 0.080 | 151 | Iraq | 0.000 | 0.000 | 0.000 |
| 36 | Iran | 0.094 | 0.101 | 0.065 | 94 | Moldova | 0.009 | 0.000 | 0.000 | 152 | Jamaica | 0.000 | 0.000 | 0.000 |
| 37 | Netherlands | 0.094 | 0.144 | 0.214 | 95 | Kyrgyzstan | 0.009 | 0.000 | 0.000 | 153 | Jordan | 0.000 | 0.000 | 0.000 |
| 38 | Tanzania | 0.092 | 0.055 | 0.067 | 96 | UK | 0.009 | 0.000 | 0.000 | 154 | Lesotho | 0.000 | 0.000 | 0.000 |
| 39 | Guatemala | 0.091 | 0.068 | 0.045 | 97 | Algeria | 0.009 | 0.001 | 0.021 | 155 | Lithuania | 0.000 | 0.000 | 0.000 |
| 40 | Belize | 0.083 | 0.096 | 0.167 | 98 | Malawi | 0.008 | 0.012 | 0.037 | 156 | Luxembourg | 0.000 | 0.000 | 0.000 |
| 41 | Fiji | 0.078 | 0.000 | 0.000 | 99 | Colombia | 0.008 | 0.000 | 0.000 | 157 | Madagascar | 0.000 | 0.000 | 0.000 |
| 42 | Myanmar | 0.077 | 0.042 | 0.200 | 100 | Zimbabwe | 0.008 | 0.000 | 0.000 | 158 | Malta | 0.000 | 0.000 | 0.000 |
| 43 | Brunei | 0.073 | 0.000 | 0.000 | 101 | Portugal | 0.008 | 0.000 | 0.000 | 159 | Mauritius | 0.000 | 0.000 | 0.000 |
| 44 | Micronesia | 0.072 | 0.000 | 0.000 | 102 | Czech Rep | 0.008 | 0.000 | 0.000 | 160 | Montenegro | 0.000 | 0.000 | 0.000 |
| 45 | Ghana | 0.069 | 0.000 | 0.000 | 103 | Japan | 0.007 | 0.008 | 0.021 | 161 | Morocco | 0.000 | 0.000 | 0.000 |
| 46 | South Africa | 0.068 | 0.100 | 0.111 | 104 | Turkmenistan | 0.007 | 0.000 | 0.000 | 162 | New Zealand | 0.000 | 0.000 | 0.000 |
| 47 | Gambia | 0.064 | 0.000 | 0.000 | 105 | Paraguay | 0.007 | 0.009 | 0.056 | 163 | Rwanda | 0.000 | 0.000 | 0.000 |
| 48 | Malaysia | 0.058 | 0.000 | 0.000 | 106 | Bangladesh | 0.007 | 0.000 | 0.000 | 164 | St Kitts | 0.000 | 0.000 | 0.000 |
| 49 | Libya | 0.056 | 0.058 | 0.136 | 107 | Austria | 0.006 | 0.000 | 0.000 | 165 | Saint Lucia | 0.000 | 0.000 | 0.000 |
| 50 | Senegal | 0.052 | 0.046 | 0.071 | 108 | Armenia | 0.006 | 0.000 | 0.000 | 166 | St Vincent | 0.000 | 0.000 | 0.000 |
| 51 | Panama | 0.049 | 0.011 | 0.077 | 109 | Switzerland | 0.005 | 0.000 | 0.000 | 167 | Samoa | 0.000 | 0.000 | 0.000 |
| 52 | Russia | 0.048 | 0.061 | 0.072 | 110 | Denmark | 0.005 | 0.000 | 0.000 | 168 | Seychelles | 0.000 | 0.000 | 0.000 |
| 53 | Laos | 0.048 | 0.000 | 0.000 | 111 | Hungary | 0.004 | 0.000 | 0.000 | 169 | South Korea | 0.000 | 0.000 | 0.000 |
| 54 | Sudan | 0.045 | 0.058 | 0.111 | 112 | Kuwait | 0.004 | 0.000 | 0.000 | 170 | Eswatini | 0.000 | 0.000 | 0.000 |
| 55 | Kosovo | 0.043 | 0.000 | 0.000 | 113 | Mauritania | 0.004 | 0.000 | 0.000 | 171 | Tonga | 0.000 | 0.000 | 0.000 |
| 56 | Uzbekistan | 0.043 | 0.058 | 0.071 | 114 | Tunisia | 0.004 | 0.000 | 0.000 | 172 | UAE | 0.000 | 0.000 | 0.000 |
| 57 | Uganda | 0.040 | 0.029 | 0.052 | 115 | Ecuador | 0.004 | 0.000 | 0.000 | 173 | Uruguay | 0.000 | 0.000 | 0.000 |
| 58 | Finland | 0.038 | 0.000 | 0.000 | 116 | Bhutan | 0.004 | 0.000 | 0.000 | | | | | |

This table provides for each country three measures of the propensity to secede: the share of the country's population in favor of secession, the share of the country's population living in regions with a majority in favor of secession, and the share of the country's regions with a majority in favor of secession. These predicted measures are based on the alternative calibration. Abbreviations used: Antigua: Antigua and Barbuda; Bosnia: Bosnia and Herzegovina; CAR: Central African Republic; Czech Rep: Czech Republic; DRC: Democratic Republic of the Congo; Dom Rep: Dominican Republic; Eq Ginea: Equatorial Guinea; Papua NG: Papua New Guinea; Rep Congo: Republic of Congo; Solomon Isl: Solomon Islands; St Kitts: Saint Kitts and Nevis; St Vincent: Saint Vincent and the Grenadines; Trinidad: Trinidad and Tobago; UAE: United Arab Emirates.

Figure C.4: Share of Population in Favor of Secession: Alternative Calibration

(a) Share of Regional Population




(b) Share of Country Population

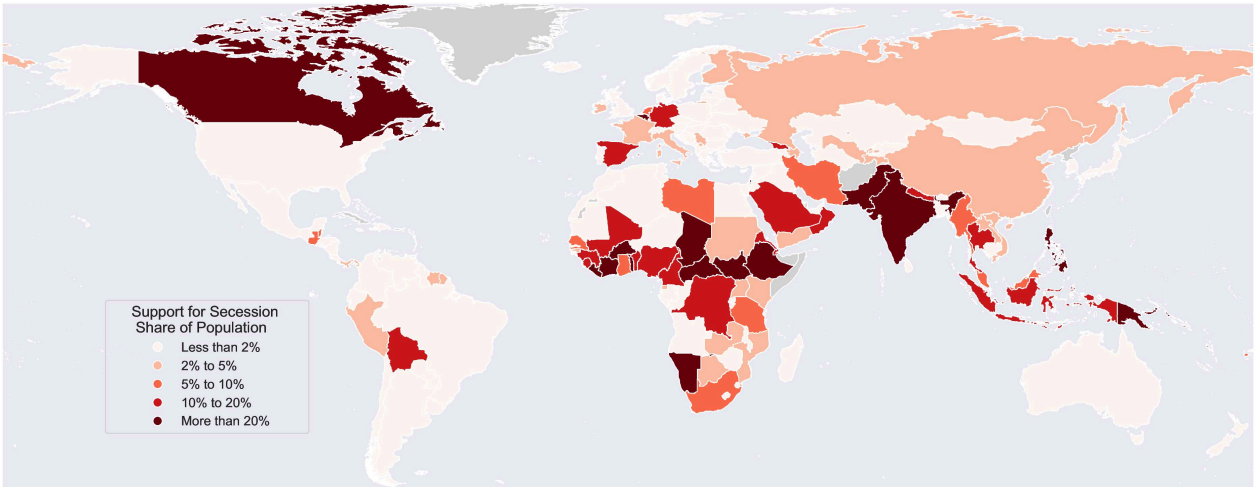



Figure depicts the model-generated share of the population in favor of secession for each subnational region (Panel (a)) and each country (Panel (b)). Results are based on the alternative calibration.

## C.5 Identity vs Income per Capita

Table C.3: Instability of Countries: Identity vs Income per Capita

| | | | Δ Pro-Secession | | | | | Δ Pro-Secession | | | | | Δ Pro-Secession | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Base-line | Iden Only | Inc Only | | | Base-line | Iden Only | Inc Only | | | Base-Line | Iden Only | Inc Only |
| 1 | Micronesia | 0.678 | 0.000 | -0.678 | 59 | Belize | 0.059 | 0.000 | -0.059 | 117 | Denmark | 0.005 | 0.000 | -0.005 |
| 2 | South Sudan | 0.632 | -0.008 | -0.632 | 60 | Uganda | 0.055 | 0.000 | -0.055 | 118 | Bolivia | 0.004 | 0.137 | -0.004 |
| 3 | Comoros | 0.415 | 0.000 | -0.415 | 61 | Serbia | 0.053 | 0.000 | -0.053 | 119 | Kuwait | 0.004 | -0.004 | -0.004 |
| 4 | Vanuatu | 0.367 | -0.067 | -0.284 | 62 | Panama | 0.052 | 0.000 | -0.052 | 120 | Sweden | 0.004 | 0.000 | -0.004 |
| 5 | Papua NG | 0.363 | 0.085 | -0.363 | 63 | Philippines | 0.052 | 0.091 | -0.052 | 121 | Tunisia | 0.004 | 0.000 | -0.004 |
| 6 | India | 0.336 | 0.136 | -0.336 | 64 | Yemen | 0.049 | -0.029 | -0.019 | 122 | Mauritania | 0.004 | 0.000 | -0.004 |
| 7 | Cameroon | 0.335 | -0.066 | -0.335 | 65 | Russia | 0.047 | 0.011 | -0.037 | 123 | Macedonia | 0.004 | 0.000 | -0.004 |
| 8 | Togo | 0.319 | 0.000 | -0.319 | 66 | Finland | 0.045 | -0.001 | -0.045 | 124 | USA | 0.004 | 0.000 | -0.004 |
| 9 | Tanzania | 0.319 | -0.005 | -0.319 | 67 | Madagascar | 0.045 | 0.000 | -0.045 | 125 | Nicaragua | 0.003 | 0.000 | -0.003 |
| 10 | Benin | 0.274 | 0.000 | -0.274 | 68 | Syria | 0.043 | 0.007 | -0.043 | 126 | Slovakia | 0.003 | 0.000 | -0.003 |
| 11 | Solomon Isl | 0.254 | 0.000 | -0.254 | 69 | Kosovo | 0.043 | 0.000 | -0.043 | 127 | Norway | 0.003 | 0.001 | -0.003 |
| 12 | Côte d'Ivoire | 0.235 | 0.000 | -0.235 | 70 | Senegal | 0.043 | 0.000 | -0.043 | 128 | Poland | 0.002 | 0.002 | -0.002 |
| 13 | Chad | 0.229 | -0.034 | -0.161 | 71 | Malawi | 0.043 | 0.000 | -0.043 | 129 | Australia | 0.002 | 0.000 | -0.002 |
| 14 | Sudan | 0.224 | -0.180 | -0.052 | 72 | Uzbekistan | 0.042 | 0.000 | -0.042 | 130 | Sri Lanka | 0.002 | 0.000 | -0.002 |
| 15 | Namibia | 0.202 | -0.127 | -0.057 | 73 | Laos | 0.042 | 0.041 | -0.042 | 131 | Slovenia | 0.002 | 0.000 | -0.002 |
| 16 | Pakistan | 0.198 | 0.033 | -0.198 | 74 | Libya | 0.041 | -0.036 | -0.005 | 132 | Latvia | 0.002 | 0.073 | -0.002 |
| 17 | Nigeria | 0.195 | 0.015 | -0.195 | 75 | France | 0.037 | -0.005 | -0.037 | 133 | Chile | 0.002 | 0.014 | -0.002 |
| 18 | DRC | 0.188 | 0.090 | -0.188 | 76 | Vietnam | 0.035 | 0.013 | -0.035 | 134 | El Salvador | 0.002 | 0.000 | -0.002 |
| 19 | Oman | 0.175 | 0.000 | -0.175 | 77 | Bhutan | 0.035 | 0.000 | -0.035 | 135 | Bulgaria | 0.002 | 0.000 | -0.002 |
| 20 | Indonesia | 0.171 | 0.056 | -0.171 | 78 | Peru | 0.034 | 0.091 | -0.034 | 136 | Iraq | 0.002 | 0.000 | -0.002 |
| 21 | CAR | 0.166 | 0.135 | -0.166 | 79 | Azerbaijan | 0.031 | 0.000 | -0.031 | 137 | Trinidad | 0.001 | 0.000 | -0.001 |
| 22 | Rep Congo | 0.159 | 0.000 | -0.159 | 80 | Suriname | 0.027 | 0.000 | -0.027 | 138 | Albania | 0.001 | 0.000 | -0.001 |
| 23 | Thailand | 0.158 | -0.138 | 0.004 | 81 | Mexico | 0.027 | 0.015 | -0.027 | 139 | Bosnia | 0.001 | 0.000 | -0.001 |
| 24 | Burkina Faso | 0.155 | 0.000 | -0.155 | 82 | Ecuador | 0.024 | 0.000 | -0.024 | 140 | Israel | 0.001 | 0.000 | -0.001 |
| 25 | Nepal | 0.154 | 0.000 | -0.154 | 83 | Guyana | 0.023 | 0.000 | -0.023 | 141 | Dom Rep | 0.001 | 0.000 | -0.001 |
| 26 | South Africa | 0.148 | 0.093 | -0.148 | 84 | Tajikistan | 0.021 | 0.000 | -0.021 | 142 | Ireland | 0.000 | 0.021 | 0.000 |
| 27 | Djibouti | 0.144 | 0.000 | -0.144 | 85 | Gambia | 0.020 | 0.000 | -0.020 | 143 | Argentina | 0.000 | 0.005 | 0.000 |
| 28 | China | 0.143 | -0.009 | -0.143 | 86 | Gabon | 0.020 | 0.000 | -0.020 | 144 | Cyprus | 0.000 | 0.000 | 0.000 |
| 29 | Zimbabwe | 0.140 | 0.000 | -0.140 | 87 | Egypt | 0.019 | -0.005 | -0.015 | 145 | UAE | 0.000 | 0.000 | 0.000 |
| 30 | Guinea | 0.140 | 0.000 | -0.140 | 88 | Brazil | 0.017 | 0.000 | -0.017 | 146 | Haiti | 0.000 | 0.000 | 0.000 |
| 31 | Eq Guinea | 0.138 | 0.000 | -0.138 | 89 | Jordan | 0.016 | 0.000 | -0.016 | 147 | Antigua | 0.000 | 0.000 | 0.000 |
| 32 | Mali | 0.135 | 0.000 | -0.135 | 90 | Lebanon | 0.015 | 0.000 | -0.015 | 148 | Bahrain | 0.000 | 0.000 | 0.000 |
| 33 | Ghana | 0.133 | 0.076 | -0.133 | 91 | Algeria | 0.014 | -0.003 | -0.013 | 149 | Belarus | 0.000 | 0.000 | 0.000 |
| 34 | Saudi Arabia | 0.131 | -0.131 | 0.004 | 92 | Honduras | 0.014 | 0.000 | -0.014 | 150 | Burundi | 0.000 | 0.000 | 0.000 |
| 35 | Bangladesh | 0.126 | -0.001 | -0.126 | 93 | Costa Rica | 0.014 | 0.000 | -0.014 | 151 | Cape Verde | 0.000 | 0.000 | 0.000 |
| 36 | Italy | 0.126 | 0.058 | -0.126 | 94 | Colombia | 0.014 | -0.004 | -0.009 | 152 | Estonia | 0.000 | 0.000 | 0.000 |
| 37 | Eritrea | 0.122 | 0.000 | -0.122 | 95 | Croatia | 0.014 | -0.003 | -0.014 | 153 | Grenada | 0.000 | 0.000 | 0.000 |
| 38 | Myanmar | 0.120 | 0.008 | -0.120 | 96 | Mongolia | 0.013 | 0.001 | -0.013 | 154 | Iceland | 0.000 | 0.000 | 0.000 |
| 39 | Guinea-Bissau | 0.116 | 0.000 | -0.116 | 97 | Cambodia | 0.012 | 0.000 | -0.012 | 155 | Jamaica | 0.000 | 0.000 | 0.000 |
| 40 | Iran | 0.115 | 0.063 | -0.115 | 98 | Hungary | 0.011 | -0.008 | -0.011 | 156 | Kazakhstan | 0.000 | 0.017 | 0.000 |
| 41 | Ethiopia | 0.114 | 0.052 | -0.114 | 99 | Romania | 0.011 | -0.005 | -0.011 | 157 | Lithuania | 0.000 | 0.000 | 0.000 |
| 42 | Germany | 0.112 | -0.035 | -0.112 | 100 | UK | 0.010 | 0.000 | -0.010 | 158 | Luxembourg | 0.000 | 0.000 | 0.000 |
| 43 | Kenya | 0.109 | -0.002 | -0.068 | 101 | Venezuela | 0.010 | 0.001 | -0.010 | 159 | Malta | 0.000 | 0.000 | 0.000 |
| 44 | Netherlands | 0.109 | 0.000 | -0.109 | 102 | Turkey | 0.010 | 0.014 | -0.010 | 160 | Mauritius | 0.000 | 0.000 | 0.000 |
| 45 | Spain | 0.108 | -0.047 | -0.108 | 103 | Czech Rep | 0.010 | -0.001 | -0.010 | 161 | Montenegro | 0.000 | 0.000 | 0.000 |
| 46 | Sierra Leone | 0.097 | 0.000 | -0.097 | 104 | Ukraine | 0.009 | 0.016 | -0.009 | 162 | Morocco | 0.000 | 0.001 | 0.000 |
| 47 | Belgium | 0.085 | 0.048 | -0.085 | 105 | Moldova | 0.009 | 0.000 | -0.009 | 163 | New Zealand | 0.000 | 0.000 | 0.000 |
| 48 | Mozambique | 0.083 | 0.111 | -0.083 | 106 | Kyrgyzstan | 0.009 | -0.009 | -0.009 | 164 | Rwanda | 0.000 | 0.000 | 0.000 |
| 49 | Georgia | 0.079 | 0.000 | -0.079 | 107 | Austria | 0.009 | 0.000 | -0.009 | 165 | St Kitts | 0.000 | 0.000 | 0.000 |
| 50 | Brunei | 0.073 | 0.000 | -0.073 | 108 | Greece | 0.009 | 0.041 | -0.009 | 166 | St Lucia | 0.000 | 0.000 | 0.000 |
| 51 | Angola | 0.073 | -0.002 | -0.073 | 109 | Paraguay | 0.008 | 0.000 | -0.008 | 167 | St Vincent | 0.000 | 0.000 | 0.000 |
| 52 | Fiji | 0.073 | 0.000 | -0.073 | 110 | Japan | 0.007 | 0.000 | -0.007 | 168 | Samoa | 0.000 | 0.000 | 0.000 |
| 53 | Malaysia | 0.070 | 0.001 | -0.070 | 111 | Turkmenistan | 0.007 | 0.000 | -0.007 | 169 | Seychelles | 0.000 | 0.000 | 0.000 |
| 54 | Liberia | 0.069 | 0.000 | -0.069 | 112 | Switzerland | 0.007 | 0.000 | -0.007 | 170 | South Korea | 0.000 | 0.000 | 0.000 |
| 55 | Zambia | 0.067 | 0.184 | -0.067 | 113 | Lesotho | 0.007 | 0.000 | -0.007 | 171 | Eswatini | 0.000 | 0.000 | 0.000 |
| 56 | Botswana | 0.066 | 0.041 | -0.066 | 114 | Portugal | 0.006 | -0.002 | -0.006 | 172 | Tonga | 0.000 | 0.000 | 0.000 |
| 57 | Guatemala | 0.065 | 0.000 | -0.065 | 115 | Armenia | 0.006 | 0.000 | -0.006 | 173 | Uruguay | 0.000 | 0.000 | 0.000 |
| 58 | Niger | 0.059 | -0.032 | -0.059 | 116 | Canada | 0.006 | 0.229 | -0.006 | | | | | |

This table reports for each country how support for secession depends on linguistic identity and on income per capita. Column 'Base' reports the share of the population that favors secession in the baseline calibration; column 'Iden' reports the change in that share if differences in income per capita are eliminated and only identity matters; column 'Inc' reports the change in that share if differences in identity are eliminated and only differences in income per capita matter. Antigua: Antigua and Barbuda; Bosnia: Bosnia and Herzegovina; CAR: Central African Republic; DRC: Democratic Republic of the Congo; Dom Rep: Dominican Republic; Papua NG: Papua New Guinea; Rep Congo: Republic of Congo; Solomon Isl: Solomon Islands; St Kitts: Saint Kitts and Nevis; St Vincent: Saint Vincent and the Grenadines; Trinidad: Trinidad and Tobago; UAE: United Arab Emirates.

## C.6 Accuracy of Model Predictions: Additional Results

Table C.4: Predicted Country Instability and Demand for Secession: Not Accounting for Regional Fixed Effects (IHS)

| | Interest in Secession | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Secessionist Activity in Country | | | | | | Number of Wikipedia Page Views (IHS) | | |
| | # Secessionist Regions (IHS) | | | # Secessionist Groups (IHS) | | | Secessionist Organization | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| # Secessionist Regions (IHS) | 0.340*** | | | 0.513*** | | | 1.580** | | |
| | (0.095) | | | (0.164) | | | (0.651) | | |
| Share of Pop. Pro-Secession (IHS) | | 2.227*** | | | 3.408** | | | 10.874* | |
| | | (0.764) | | | (1.337) | | | (5.569) | |
| At least 1 Region Pro-Secession | | | 0.592*** | | | 0.910*** | | | 2.601** |
| | | | (0.163) | | | (0.265) | | | (1.083) |
| WB Region FE | No | No | No | No | No | No | No | No | No |
| # Regions | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.27 | 0.25 | 0.27 | 0.19 | 0.18 | 0.20 | 0.10 | 0.10 | 0.10 |
| Observations | 173 | 173 | 173 | 173 | 173 | 173 | 173 | 173 | 173 |

Notes: Table reports results from regressions of different measures of secessionist activity from Wikipedia on model-generated measures of the demand for secession. It does so at the level of countries, and it includes the number of subnational regions as an additional control. Compared to Table 1 Panel A in the main text, it does not include World Bank region fixed effects. IHS denotes variables have been transformed using the inverse hyperbolic sine transformation. Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C.5: Predicted Country Instability and Demand for Secession: Dropping Regions

Panel A: Number of Secessionist Regions by Country

| | Secessionist Activity in Country - Number of Secessionist Regions | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Full Sample | No EAP | No ECA | No LAC | No MENA | No NAM | No SA | No SSA |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| # Secessionist Regions (IHS) | 0.344*** | 0.333** | 0.262** | 0.375*** | 0.353*** | 0.344*** | 0.353*** | 0.381*** |
| | (0.110) | (0.134) | (0.117) | (0.116) | (0.117) | (0.110) | (0.116) | (0.135) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| # Regions (IHS) | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.31 | 0.28 | 0.39 | 0.31 | 0.30 | 0.28 | 0.32 | 0.34 |
| Observations | 173 | 151 | 124 | 144 | 154 | 171 | 167 | 127 |

Panel B: Number of Secessionist Regions by Country

| | Secessionist Activity in Country - Number of Secessionist Groups | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Full Sample | No EAP | No ECA | No LAC | No MENA | No NAM | No SA | No SSA |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| # Secessionist Regions (IHS) | 0.542*** | 0.572** | 0.384* | 0.587*** | 0.547*** | 0.541*** | 0.553*** | 0.611*** |
| | (0.185) | (0.221) | (0.196) | (0.195) | (0.194) | (0.185) | (0.195) | (0.223) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| # Regions (IHS) | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.26 | 0.23 | 0.34 | 0.22 | 0.24 | 0.23 | 0.25 | 0.30 |
| Observations | 173 | 151 | 124 | 144 | 154 | 171 | 167 | 127 |

Notes: Table reports results from regressing the number of secessionist regions (Panel A) or the number of secessionist groups (Panel B) on the model-predicted number of secessionist regions. It does so at the level of countries, and it includes World Bank region fixed effects and the number of subnational regions as additional controls. Column (1) includes the full sample, whereas columns (2)-(8) each drop one of the World Bank regions, where EAP refers to East Asia and the Pacific, ECA refers to Europe and Central Asia, LAC refers to Latin America and the Caribbean, MENA refers to Middle East and North Africa, NAM refers to North America, SA refers to South Asia, and SSA refers to sub-Saharan Africa. IHS denotes variables have been transformed using the inverse hyperbolic sine transformation. Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C.6: Predicted Secession and Contemporary Secessionist Activity: Sambanis et al.

Panel A: Sambanis et al.

| | Secessionist Activity in Country | | | # Secessionist Groups in Country (IHS) | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| # Secessionist Regions (IHS) | 0.136*** | | | 0.858*** | | |
| | (0.041) | | | (0.189) | | |
| Share of Pop. Pro-Secession (IHS) | | 1.201*** | | | 6.005*** | |
| | | (0.417) | | | (1.971) | |
| At least 1 Region Pro-Secession | | | 0.236*** | | | 1.378*** |
| | | | (0.071) | | | (0.315) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes |
| # Regions (IHS) | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.13 | 0.15 | 0.14 | 0.23 | 0.21 | 0.22 |
| Observations | 173 | 173 | 173 | 173 | 173 | 173 |

Panel B: Sambanis et al. & Wikipedia

| | Secessionist Activity in Country | | | # Secessionist Groups in Country (IHS) | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| # Secessionist Regions (IHS) | 0.157*** | | | 0.846*** | | |
| | (0.046) | | | (0.193) | | |
| Share of Pop. Pro-Secession (IHS) | | 1.365*** | | | 6.233*** | |
| | | (0.499) | | | (2.042) | |
| At least 1 Region Pro-Secession | | | 0.304*** | | | 1.478*** |
| | | | (0.080) | | | (0.314) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes |
| # Regions (IHS) | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.12 | 0.13 | 0.14 | 0.24 | 0.23 | 0.25 |
| Observations | 173 | 173 | 173 | 173 | 173 | 173 |

Notes: Table reports results from regressions of different measures of secessionist activity on model-generated measures of the demand for secession. It does so at the level of countries, and it includes World Bank region fixed effects and the number of subnational regions as additional controls. As measures of secessionist activity, Panel A uses the presence and the number of self-determination movements in Sambanis et al. (2018) whereas Panel B defines the presence of secessionist groups as having a secessionist group in either our Wikipedia data or Sambanis et al. (2018) and it defines the number of secessionist groups as the maximum number of groups in either our Wikipedia data or Sambanis et al. (2018). IHS denotes variables have been transformed using the inverse hyperbolic sine transformation. Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

## Table C.7: Predicted Demand for Secession and Fragile State Index Subcomponents

| | Fragile State Index Subcomponents (2006-2021) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | C1 | C2 | C3 | E1 | E2 | E3 | P1 | P2 | P3 | S1 | S2 | X1 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Share of Pop. Pro-Secession (IHS) | 4.830*** | 3.355** | 3.307 | 2.899*** | 3.819*** | 2.262 | 2.924** | 4.679*** | 1.633 | 3.057*** | 3.063* | 2.654* |
| | (1.667) | (1.615) | (2.017) | (1.023) | (1.076) | (1.453) | (1.474) | (1.279) | (1.926) | (1.043) | (1.830) | (1.493) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.33 | 0.21 | 0.17 | 0.40 | 0.50 | 0.39 | 0.27 | 0.59 | 0.34 | 0.62 | 0.33 | 0.31 |
| Observations | 167 | 167 | 167 | 167 | 167 | 167 | 167 | 167 | 167 | 167 | 167 | 167 |

Notes: Table reports regressions of subcomponents of the state fragility index on the model-predicted share of the population that is pro-secession. Regressions are run at the level of countries, and include World Bank region fixed effects as additional controls. Each column shows one subcomponent of the Fragile State Index. C1: Security Apparatus, C2: Factionalized Elites, C3: Group Grievance, E1: Economic Decline, E2: Uneven Economic Development, E3: Human Flight and Brain Drain, P1: State Legitimacy, P2: Public Services, P3: Human Rights and Rule of Law, S1: Demographic Pressures, S2: Refugees and IDPs, X1: External Intervention. IHS denotes variables have been transformed using the inverse hyperbolic sine transformation. Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

## Table C.8: Predicted Demand for Secession and Regional Authority Index Subcomponents

| | Regional Authority Index (1950-2016) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Self-Rule | | | | | | Shared-Rule | | | | | |
| | SE1 | SE2 | SE3 | SE4 | SE5 | SE | SH1 | SH2 | SH3 | SH4 | SH5 | SH |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Share of Pop. | 9.442*** | 6.746*** | 4.700* | 2.185 | 12.375*** | 35.379** | 2.316* | 1.186 | 1.154 | 0.054 | 7.272** | 11.967 |
| Pro-Secession (IHS) | (3.500) | (2.427) | (2.590) | (1.932) | (4.139) | (13.549) | (1.372) | (1.124) | (1.073) | (0.656) | (3.641) | (7.306) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.30 | 0.21 | 0.22 | 0.19 | 0.27 | 0.25 | 0.02 | 0.19 | 0.01 | -0.02 | 0.17 | 0.08 |
| Observations | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 |

Notes: Table reports regressions of subcomponents of the regional authority index on the model-predicted share of the population that is pro-secession. Regressions are run at the level of countries, and include World Bank region fixed effects as additional controls. Each column shows one subcomponent of the regional authority index. First 6 columns contain subcomponents measuring the level of self-rule within regions, the last 6 columns measure the level of shared power between regions and central governments. SE1: The extent to which a regional government is autonomous rather than deconcentrated. SE2: The range of policies for which a regional government is responsible. SE3: The extent to which a regional government can independently tax its population. SE4: The extent to which a regional government can borrow. SE5: The extent to which a region has an independent legislature and executive, which is the sum of assembly and executive. SE: The authority exercised by a regional government over those who live in the region, which is the sum of SE1-SE5. SH1:The extent to which regional representatives codetermine national legislation. SH2: The extent to which a regional government codetermines national policy in intergovernmental meetings. SH3: The extent to which regional representatives codetermine the distribution of national tax revenues. SH4: The extent to which a regional government codetermines subnational and national borrowing constraints. SH5: The extent to which regional representatives codetermine constitutional change. SH: The authority exercised by a regional government or its representatives in the country as a whole, which is the sum of SH1-SH5. IHS denotes variables have been transformed using the inverse hyperbolic sine transformation. Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

## C.7 Accuracy of Model Predictions: Log Transformation

Table C.9: Predicted Demand for Secession and Contemporary Secessionist Activity (Logs)

| Panel A: Country-Level Analysis | Interest in Secession | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Secessionist Activity in Country | | | | | | Log[Number of Wikipedia Page Views] | | |
| | Log[1+# Secessionist Regions] | | | Log[1+# Secessionist Groups] | | | Secessionist Organization | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Log[1 + # Secessionist Regions] | 0.344*** | | | 0.585*** | | | 2.144** | | |
| | (0.113) | | | (0.202) | | | (0.894) | | |
| Log[1 + Share Pop. Pro-Secession] | | 2.272*** | | | 3.966*** | | | 15.107** | |
| | | (0.788) | | | (1.457) | | | (6.957) | |
| At least 1 Region Pro-Secession | | | 0.478*** | | | 0.831*** | | | 2.847** |
| | | | (0.142) | | | (0.247) | | | (1.112) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Log[1+# Regions] | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.33 | 0.33 | 0.34 | 0.27 | 0.27 | 0.28 | 0.18 | 0.18 | 0.18 |
| Observations | 173 | 173 | 173 | 173 | 173 | 173 | 173 | 173 | 173 |

| Panel B: Regional-Level Analysis | Interest in Secession | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Secessionist Activity in Region | | | | | | Log[Total Number of Wikipedia Page Views] | | |
| | Secessionist Region | | | Log[1+# Secessionist Groups in Region] | | | Secessionist Organization | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Log[1 + Sh. Pop. Pro-Secession] | 0.301*** | | | 0.459*** | | | 2.378*** | | |
| | (0.070) | | | (0.124) | | | (0.777) | | |
| At least 10%+ Pro-Secession | | 0.125*** | | | 0.188*** | | | 1.035*** | |
| | | (0.028) | | | (0.049) | | | (0.328) | |
| At least 50% Pro-Secession | | | 0.133*** | | | 0.223*** | | | 1.102*** |
| | | | (0.038) | | | (0.080) | | | (0.414) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.21 | 0.21 | 0.21 | 0.24 | 0.24 | 0.24 | 0.22 | 0.22 | 0.21 |
| Observations | 3003 | 3003 | 3003 | 3003 | 3003 | 3003 | 3003 | 3003 | 3003 |

Notes: Table reports results from regressions of different measures of secessionist activity from Wikipedia on model-generated measures of the demand for secession. Panel A does this at the level of countries, and includes World Bank region fixed effects and the number of subnational regions as additional controls, whereas Panel B does this at the level of subnational regions and includes country fixed effects. Compared to Table 1 in the main text, variables have been transformed using logarithms. Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C.10: Predicted Secession and Contemporary Secessionist Activity: Sambanis et al. (Logs)

Panel A: Sambanis et al.

| | Secessionist Activity in Country | | | Log[1+# Secessionist Groups in Country] | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Log[1 + # Secessionist Regions] | 0.197*** | | | 0.961*** | | |
| | (0.059) | | | (0.208) | | |
| Log[1 + Share Pop. Pro-Secession] | | 1.670*** | | | 6.334*** | |
| | | (0.526) | | | (1.869) | |
| At least 1 Region Pro-Secession | | | 0.267*** | | | 1.187*** |
| | | | (0.081) | | | (0.268) |
| Log[1+# Regions] | Yes | Yes | Yes | Yes | Yes | Yes |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.13 | 0.16 | 0.14 | 0.24 | 0.23 | 0.23 |
| Observations | 173 | 173 | 173 | 173 | 173 | 173 |

Panel B: Sambanis et al. & Wikipedia

| | Secessionist Activity in Country | | | Log[1+# Secessionist Groups in Country] | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Log[1 + # Secessionist Regions] | 0.201*** | | | 0.943*** | | |
| | (0.059) | | | (0.214) | | |
| Log[1 + Share Pop. Pro-Secession] | | 1.682*** | | | 6.591*** | |
| | | (0.556) | | | (1.931) | |
| At least 1 Region Pro-Secession | | | 0.303*** | | | 1.266*** |
| | | | (0.080) | | | (0.268) |
| Log[1+# Regions] | Yes | Yes | Yes | Yes | Yes | Yes |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.12 | 0.14 | 0.14 | 0.25 | 0.26 | 0.26 |
| Observations | 173 | 173 | 173 | 173 | 173 | 173 |

Notes: Table reports results from regressions of different measures of secessionist activity on model-generated measures of the demand for secession. It does so at the level of countries, and it includes World Bank region fixed effects and the number of subnational regions as additional controls. As measures of secessionist activity, Panel A uses the presence and the number of self-determination movements in Sambanis et al. (2018) whereas Panel B defines the presence of secessionist groups as having a secessionist group in either our Wikipedia data or Sambanis et al. (2018) and it defines the number of secessionist groups as the maximum number of groups in either our Wikipedia data or Sambanis et al. (2018). Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C.11: Predicted Demand for Secession and State Fragility, Regional Autonomy, and Conflict (Logs)

| Panel A: Institutional | Fragile State Index (2006-2021) | | | | | | Regional Authority Index (1950-2016) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Rank | | | Index | | | Total | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Log[1 + # Secessionist Regions] | 19.343*** | | | 8.038*** | | | 3.784** | | |
| | (5.346) | | | (2.393) | | | (1.567) | | |
| Log[1 + Share Pop. Pro-Secession] | | 97.012*** | | | 46.765*** | | | 53.527** | |
| | | (36.300) | | | (15.742) | | | (21.725) | |
| At least 1 Region Pro-Secession | | | 24.114*** | | | 9.749*** | | | 5.208** |
| | | | (7.061) | | | (3.150) | | | (2.362) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.45 | 0.43 | 0.45 | 0.43 | 0.42 | 0.43 | 0.19 | 0.22 | 0.19 |
| Observations | 167 | 167 | 167 | 167 | 167 | 167 | 89 | 89 | 89 |

| Panel B: Conflict | Intensity of Conflict (1997-2020) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Log[1 + # Deaths] | | | Log[1 + # Events] | | | Log[1 + # Years] | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Log[1 + # Secessionist Regions] | 1.896*** | | | 1.333*** | | | 0.296** | | |
| | (0.404) | | | (0.437) | | | (0.116) | | |
| Log[1 + Share Pop. Pro-Secession] | | 4.386 | | | 0.113 | | | -0.060 | |
| | | (3.676) | | | (3.671) | | | (1.092) | |
| At least 1 Region Pro-Secession | | | 2.268*** | | | 1.486** | | | 0.425** |
| | | | (0.553) | | | (0.573) | | | (0.163) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.45 | 0.38 | 0.44 | 0.24 | 0.19 | 0.22 | 0.66 | 0.64 | 0.66 |
| Observations | 173 | 173 | 173 | 173 | 173 | 173 | 173 | 173 | 173 |

Notes: Table reports results from regressions of measures of state fragility, regional autonomy, and conflict on model-generated measures of the demand for secession. Regressions are run at the level of countries, and include World Bank region fixed effects as additional controls. Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C.12: Predicted Demand for Secession and Fragile State Index Subcomponents (Logs)

| | Fragile State Index Subcomponents (2006-2021) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | C1 | C2 | C3 | E1 | E2 | E3 | P1 | P2 | P3 | S1 | S2 | X1 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Log[1 + Share Pop. Pro-Secession] | 5.940*** | 4.164** | 4.173* | 3.354*** | 4.630*** | 2.587 | 3.612** | 5.627*** | 2.174 | 3.729*** | 3.840* | 2.918 |
| | (1.908) | (1.892) | (2.319) | (1.253) | (1.279) | (1.703) | (1.749) | (1.504) | (2.206) | (1.256) | (2.128) | (1.824) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Mean Dependent | 5.71 | 6.45 | 6.24 | 5.65 | 5.88 | 5.67 | 6.05 | 5.50 | 5.75 | 5.76 | 5.32 | 5.75 |
| Adjusted-$R^2$ | 0.34 | 0.21 | 0.17 | 0.40 | 0.50 | 0.39 | 0.27 | 0.59 | 0.34 | 0.62 | 0.34 | 0.31 |
| Observations | 167 | 167 | 167 | 167 | 167 | 167 | 167 | 167 | 167 | 167 | 167 | 167 |

Notes: Table reports regressions of subcomponents of the state fragility index on the model-predicted share of the population that is pro-secession. Regressions are run at the level of countries, and include World Bank region fixed effects as additional controls. Each column shows one subcomponent of the Fragile State Index. C1: Security Apparatus, C2: Factionalized Elites, C3: Group Grievance, E1: Economic Decline, E2: Uneven Economic Development, E3: Human Flight and Brain Drain, P1: State Legitimacy, P2: Public Services, P3: Human Rights and Rule of Law, S1: Demographic Pressures, S2: Refugees and IDPs, X1: External Intervention. Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C.13: Predicted Demand for Secession and Regional Authority Index Subcomponents (Logs)

| | Regional Authority Index (1950-2016) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Self-Rule | | | | | | Shared-Rule | | | | | |
| | SE1 | SE2 | SE3 | SE4 | SE5 | SE | SH1 | SH2 | SH3 | SH4 | SH5 | SH |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Log[1 + Share Pop. | 10.715*** | 7.500*** | 5.149* | 2.514 | 14.025*** | 39.820*** | 2.621* | 1.342 | 1.376 | 0.091 | 8.292** | 13.706* |
| Pro-Secession] | (3.729) | (2.674) | (2.819) | (2.141) | (4.445) | (14.625) | (1.502) | (1.251) | (1.227) | (0.754) | (3.958) | (8.054) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Mean Dependent | 2.21 | 1.70 | 1.06 | 1.02 | 3.04 | 9.03 | 0.38 | 0.31 | 0.28 | 0.18 | 0.85 | 2.00 |
| Adjusted-$R^2$ | 0.31 | 0.21 | 0.22 | 0.19 | 0.28 | 0.25 | 0.02 | 0.19 | 0.01 | -0.02 | 0.17 | 0.09 |
| Observations | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 |

Notes: Table reports regressions of subcomponents of the regional authority index on the model-predicted share of the population that is pro-secession. Regressions are run at the level of countries, and include World Bank region fixed effects as additional controls. Each column shows one subcomponent of the regional authority index. First 6 columns contain subcomponents measuring the level of self-rule within regions, the last 6 columns measure the level of shared power between regions and central governments. SE1: The extent to which a regional government is autonomous rather than deconcentrated. SE2: The range of policies for which a regional government is responsible. SE3: The extent to which a regional government can independently tax its population. SE4: The extent to which a regional government can borrow. SE5: The extent to which a region has an independent legislature and executive, which is the sum of assembly and executive. SE: The authority exercised by a regional government over those who live in the region, which is the sum of SE1-SE5. SH1:The extent to which regional representatives codetermine national legislation. SH2: The extent to which a regional government codetermines national policy in intergovernmental meetings. SH3: The extent to which regional representatives codetermine the distribution of national tax revenues. SH4: The extent to which a regional government codetermines subnational and national borrowing constraints. SH5: The extent to which regional representatives codetermine constitutional change. SH: The authority exercised by a regional government or its representatives in the country as a whole, which is the sum of SH1-SH5. Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

## C.8 Accuracy of Model Predictions: Alternative Calibration

Table C.14: Predicted Secessionism and Contemporary Secessionist Activity (Alternative Calibration)

| Panel A: Country-Level Analysis | Interest in Secession | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Secessionist Activity in Country | | | | | | Number of Wikipedia Page Views (IHS) | | |
| | # Secessionist Regions (IHS) | | | # Secessionist Groups (IHS) | | | Secessionist Organization | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| # Secessionist Regions (IHS) | 0.337*** | | | 0.513*** | | | 1.892** | | |
| | (0.113) | | | (0.193) | | | (0.766) | | |
| Share of Pop. Pro-Secession (IHS) | | 1.639** | | | 2.647** | | | 11.651** | |
| | | (0.679) | | | (1.275) | | | (5.562) | |
| At least 1 Region Pro-Secession | | | 0.532*** | | | 0.820*** | | | 3.180*** |
| | | | (0.171) | | | (0.287) | | | (1.146) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| # Regions | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.31 | 0.29 | 0.31 | 0.25 | 0.24 | 0.26 | 0.18 | 0.18 | 0.19 |
| Observations | 173 | 173 | 173 | 173 | 173 | 173 | 173 | 173 | 173 |

| Panel B: Regional-Level Analysis | Interest in Secession | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Secessionist Activity in Region | | | | | | Total Number of Wikipedia Page Views (IHS) | | |
| | Secessionist Region | | | # Secessionist Groups in Region (IHS) | | | Secessionist Organization | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Sh. Pop. Pro-Secession (IHS) | 0.223*** | | | 0.462*** | | | 2.463*** | | |
| | (0.053) | | | (0.109) | | | (0.612) | | |
| At least 10%+ Pro-Secession | | 0.145*** | | | 0.282*** | | | 1.454*** | |
| | | (0.026) | | | (0.051) | | | (0.262) | |
| At least 50% Pro-Secession | | | 0.116*** | | | 0.254*** | | | 1.421*** |
| | | | (0.037) | | | (0.083) | | | (0.494) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.21 | 0.22 | 0.21 | 0.24 | 0.24 | 0.24 | 0.22 | 0.22 | 0.21 |
| Observations | 3003 | 3003 | 3003 | 3003 | 3003 | 3003 | 3003 | 3003 | 3003 |

Notes: Table reports results from regressions of different measures of secessionist activity from Wikipedia on model-generated measures of the demand for secession. The model-generated measures are based on the alternative calibration. Panel A does this at the level of countries, and includes World Bank region fixed effects and the number of subnational regions as additional controls, whereas Panel B does this at the level of subnational regions and includes country fixed effects. IHS denotes variables have been transformed using the inverse hyperbolic sine transformation. Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C.15: Predicted Secession and Contemporary Secessionist Activity: Sambanis et al. (Alternative Calibration)

Panel A: Sambanis et al.

| | Secessionist Activity in Country | | | # Secessionist Groups in Country (IHS) | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| # Secessionist Regions (IHS) | 0.083** | | | 0.652*** | | |
| | (0.042) | | | (0.201) | | |
| Share of Pop. Pro-Secession (IHS) | | 0.689*** | | | 4.190*** | |
| | | (0.261) | | | (1.309) | |
| At least 1 Region Pro-Secession | | | 0.153** | | | 0.995*** |
| | | | (0.069) | | | (0.316) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes |
| # Regions (IHS) | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.10 | 0.11 | 0.11 | 0.19 | 0.18 | 0.19 |
| Observations | 173 | 173 | 173 | 173 | 173 | 173 |

Panel B: Sambanis et al. & Wikipedia

| | Secessionist Activity in Country | | | # Secessionist Groups in Country (IHS) | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| # Secessionist Regions (IHS) | 0.117** | | | 0.723*** | | |
| | (0.047) | | | (0.200) | | |
| Share of Pop. Pro-Secession (IHS) | | 0.793*** | | | 4.437*** | |
| | | (0.302) | | | (1.333) | |
| At least 1 Region Pro-Secession | | | 0.220*** | | | 1.179*** |
| | | | (0.076) | | | (0.310) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes |
| # Regions (IHS) | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.09 | 0.09 | 0.11 | 0.22 | 0.20 | 0.22 |
| Observations | 173 | 173 | 173 | 173 | 173 | 173 |

Notes: Table reports results from regressions of different measures of secessionist activity on model-generated measures of the demand for secession. The model-generated measures are based on the alternative calibration. It does so at the level of countries, and it includes World Bank region fixed effects and the number of subnational regions as additional controls. As measures of secessionist activity, Panel A uses the presence and the number of self-determination movements in Sambanis et al. (2018) whereas Panel B defines the presence of secessionist groups as having a secessionist group in either our Wikipedia data or Sambanis et al. (2018) and it defines the number of secessionist groups as the maximum number of groups in either our Wikipedia data or Sambanis et al. (2018). IHS denotes variables have been transformed using the inverse hyperbolic sine transformation. Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C.16: Predicted Demand for Secession and State Fragility, Regional Autonomy, and Conflict (Alternative Calibration)

| Panel A: Institutional | Fragile State Index (2006-2021) | | | | | | Regional Authority Index (1950-2016) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Rank | | | Index | | | Total | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| # Secessionist Regions (IHS) | 16.611*** | | | 7.119*** | | | 3.776** | | |
| | (3.758) | | | (1.677) | | | (1.438) | | |
| Share of Pop. Pro-Secession (IHS) | | 65.771** | | | 26.285* | | | 31.794*** | |
| | | (28.634) | | | (13.684) | | | (10.654) | |
| At least 1 Region Pro-Secession | | | 24.139*** | | | 10.263*** | | | 6.766** |
| | | | (6.483) | | | (2.936) | | | (2.864) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.47 | 0.42 | 0.45 | 0.45 | 0.41 | 0.44 | 0.15 | 0.19 | 0.16 |
| Observations | 167 | 167 | 167 | 167 | 167 | 167 | 89 | 89 | 89 |

| Panel B: Conflict | Intensity of Conflict (1997-2020) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | # Deaths (IHS) | | | # Events (IHS) | | | # Years (IHS) | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| # Secessionist Regions (IHS) | 1.773*** | | | 1.299*** | | | 0.342*** | | |
| | (0.330) | | | (0.350) | | | (0.104) | | |
| Share of Pop. Pro-Secession (IHS) | | 6.405** | | | 2.833 | | | 5.810 | |
| | | (2.782) | | | (2.935) | | | (3.514) | |
| At least 1 Region Pro-Secession | | | 2.446*** | | | 1.833*** | | | 0.587*** |
| | | | (0.541) | | | (0.553) | | | (0.166) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.48 | 0.40 | 0.45 | 0.26 | 0.20 | 0.24 | 0.65 | 0.75 | 0.65 |
| Observations | 173 | 173 | 173 | 173 | 173 | 173 | 173 | 173 | 173 |

Notes: Table reports results from regressions of measures of state fragility, regional autonomy, and conflict on model-generated measures of the demand for secession. The model-generated measures are based on the alternative calibration. Regressions are run at the level of countries, and include World Bank region fixed effects as additional controls. IHS denotes variables have been transformed using the inverse hyperbolic sine transformation. Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C.17: Predicted Secession and Fragile State Index Subcomponents (Alternative Calibration)

| | Fragile State Index Subcomponents (2006-2021) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | C1 | C2 | C3 | E1 | E2 | E3 | P1 | P2 | P3 | S1 | S2 | X1 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Share of Pop. Pro-Secession (IHS) | 4.054*** | 2.633** | 3.054** | 1.383 | 2.336** | 0.500 | 1.221 | 3.020** | 1.421 | 2.255** | 2.912* | 1.480 |
| | (1.522) | (1.329) | (1.291) | (0.876) | (1.090) | (1.149) | (1.470) | (1.336) | (1.377) | (1.018) | (1.563) | (1.459) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.33 | 0.21 | 0.17 | 0.39 | 0.48 | 0.38 | 0.26 | 0.58 | 0.34 | 0.62 | 0.34 | 0.30 |
| Observations | 167 | 167 | 167 | 167 | 167 | 167 | 167 | 167 | 167 | 167 | 167 | 167 |

Notes: Table reports regressions of subcomponents of the state fragility index on the model-predicted share of the population that is pro-secession. The model-predicted share is based on the alternative calibration. Regressions are run at the level of countries, and include World Bank region fixed effects as additional controls. Each column shows one subcomponent of the Fragile State Index. C1: Security Apparatus, C2: Factionalized Elites, C3: Group Grievance, E1: Economic Decline, E2: Uneven Economic Development, E3: Human Flight and Brain Drain, P1: State Legitimacy, P2: Public Services, P3: Human Rights and Rule of Law, S1: Demographic Pressures, S2: Refugees and IDPs, X1: External Intervention. IHS denotes variables have been transformed using the inverse hyperbolic sine transformation. Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C.18: Predicted Secession and Regional Autonomy Index Subcomponents (Alternative Calibration)

| | Regional Autonomy Index (1950-2016) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Self-Rule | | | | | | Shared-Rule | | | | | |
| | SE1 | SE2 | SE3 | SE4 | SE5 | SE | SH1 | SH2 | SH3 | SH4 | SH5 | SH |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Share of Pop. | 3.731*** | 3.318*** | 2.876*** | 0.644 | 4.733*** | 15.294*** | 1.194** | 0.659 | 1.151 | 0.331 | 4.588** | 7.916** |
| Pro-Secession (IHS) | (1.157) | (0.876) | (1.025) | (0.613) | (1.257) | (4.562) | (0.543) | (0.427) | (0.767) | (0.345) | (1.838) | (3.689) |
| WB Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted-$R^2$ | 0.23 | 0.20 | 0.24 | 0.18 | 0.21 | 0.22 | 0.02 | 0.20 | 0.06 | -0.01 | 0.22 | 0.12 |
| Observations | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 |

Notes: Table reports regressions of subcomponents of the regional authority index on the model-predicted share of the population that is pro-secession. The model-predicted share is based on the alternative calibration. Regressions are run at the level of countries, and include World Bank region fixed effects as additional controls. Each column shows one subcomponent of the regional authority index. First 6 columns contain subcomponents measuring the level of self-rule within regions, the last 6 columns measure the level of shared power between regions and central governments. SE1: The extent to which a regional government is autonomous rather than deconcentrated. SE2: The range of policies for which a regional government is responsible. SE3: The extent to which a regional government can independently tax its population. SE4: The extent to which a regional government can borrow. SE5: The extent to which a region has an independent legislature and executive, which is the sum of assembly and executive. SE: The authority exercised by a regional government over those who live in the region, which is the sum of SE1-SE5. SH1:The extent to which regional representatives codetermine national legislation. SH2: The extent to which a regional government codetermines national policy in intergovernmental meetings. SH3: The extent to which regional representatives codetermine the distribution of national tax revenues. SH4: The extent to which a regional government codetermines subnational and national borrowing constraints. SH5: The extent to which regional representatives codetermine constitutional change. SH: The authority exercised by a regional government or its representatives in the country as a whole, which is the sum of SH1-SH5. IHS denotes variables have been transformed using the inverse hyperbolic sine transformation. Heteroskedasticity robust standard error estimates clustered at the country-level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Figure C.5: Predicted Secession and Conflict Intensity (Deaths) across Countries (Alternative Calibration)

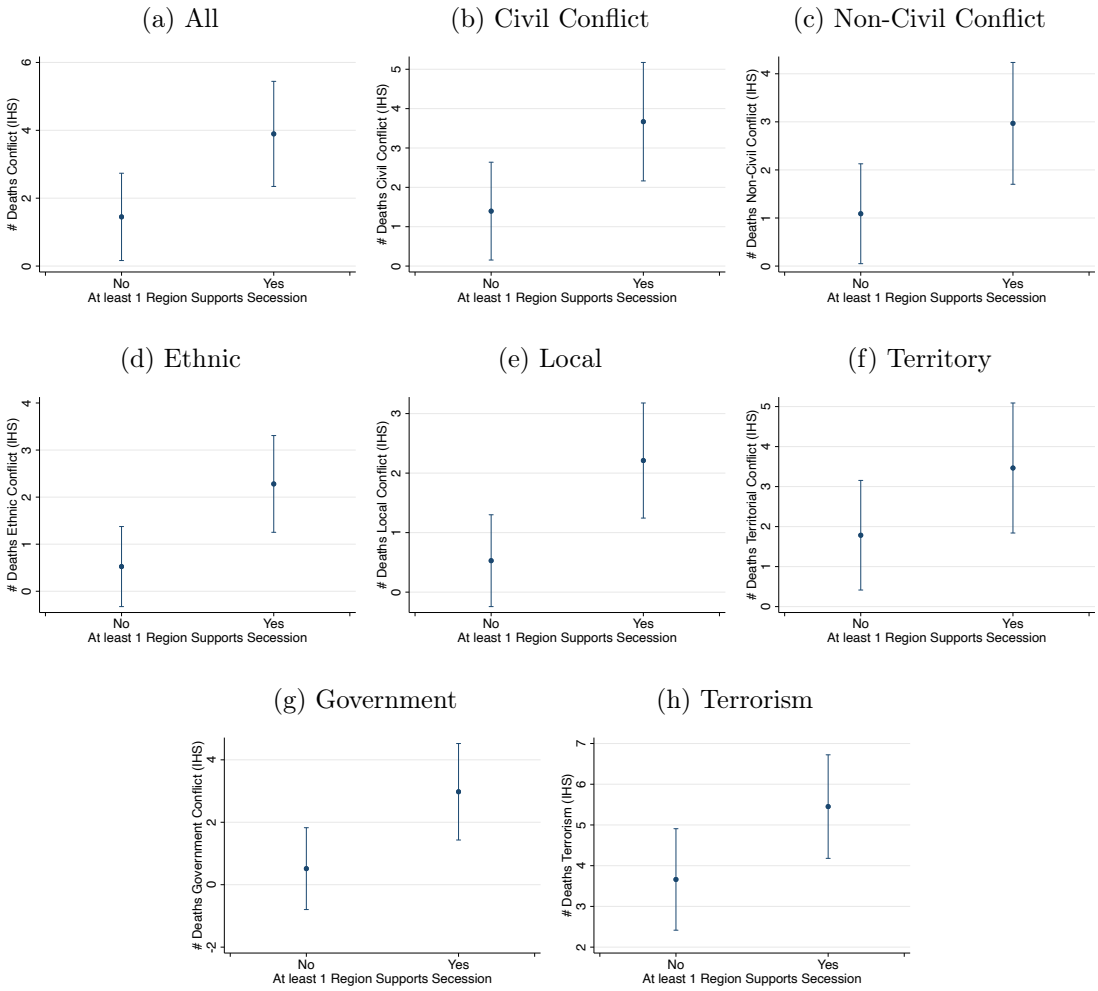



Figure shows the association between our model-based predicted demand for secession as measured by a country having at least 1 region in favor of secession and the number of deaths by type and source of conflict. The model-based predicted demand is based on the alternative calibration. IHS denotes variables have been transformed using the inverse hyperbolic sine transformation.