Globalization and Political Structure*

Gino Gancia, Giacomo A. M. Ponzetto and Jaume Ventura

July 2018

Abstract

This paper develops a theoretical framework to study the interaction between globalization and political structure. We show that political structure adapts to expanding trade opportunities in a non-monotonic way. Borders hamper trade. In its early stages, the political response to globalization consists of removing borders by increasing country size. In its later stages, however, the political response to globalization is to remove borders by creating international unions, and this leads to a reduction in country size. Moreover, negotiation replaces war as a tool to ensure market access. These predictions are consistent with historical evidence on trade, territorial expansion and war.

JEL Classification: D71, F15, F55, H77, O57

Keywords: Globalization, political structure, size of countries, international unions.

*Gino Gancia: Queen Mary University of London and CREI. Giacomo Ponzetto: CREI, Universitat Pompeu Fabra, IPEG and Barcelona GSE. Jaume Ventura: CREI, Universitat Pompeu Fabra and Barcelona GSE. E-mail: ggancia@crei.cat, gponzetto@crei.cat, jventura@crei.cat. We thank Janko Heineken, Marta Santamaria and Jagdish Tripathy for excellent research assistance. We are grateful for helpful comments by Giovanni Andreottola, Pablo Beramendi, Graziella Bertocchi, Alessandra Bonfiglioli and conference and seminar participants at the Barcelona GSE Summer Forum, Bocconi, Bologna, Bolzano, Cambridge, CEPR ERWIT, Chicago Harris, CREI, Edinburgh, EIEF, ESSIM, EUI, Georgetown, Humboldt, IEB, LSE, LUISS, Namur, NBER Summer Institute, Nottingham Ningbo, Norwegian Business School, NYU, Paris School of Economics, Princeton, QMUL, Queens College, SED, Stanford GSB, Transpyrenean Macro Workshop, Vanderbilt, and Zurich. We acknowledge financial support from the Centre de Recerca en Economia Internacional (CREI), through the European Research Council (ERC), under the European Union's Horizon 2020 Programme, Grant Agreements 693512 ("Globalization, Economic Policy and Political Structure") and 714905 ("Citizens, Institutions and Globalization"), the Spanish Ministry of Economy and Competitiveness (grants RYC-2013-13838 and SEV-2015-0563), the Generalitat de Catalunya (CERCA program and grant 2014 SGR 830) and the Institut d'Estudis Autonòmics.

1 Introduction

Since the Industrial Revolution, the cost of distance has been falling dramatically thanks to a stream of major technological innovations, like the railroad, the steamship, the telegraph, the jet engine, containerization and most recently the internet. Such technological progress has fundamentally transformed worldwide economic geography and fuelled a continuous expansion in the size of markets. In 1820, international trade was very modest: about 2% of world output. Over the following century, international trade grew more than four-fold to 8%. This first wave of globalization was cut short in the interwar period, which saw international trade decline to about 5%. After World War II, a second wave of globalization started and still continues today. By 2010, international trade had reached unprecedented levels, surpassing 20% of world output.

These two ages of globalization, however, saw political geography evolve in opposite directions. In the nineteenth century, economic and political integration proceeded together. Sovereign states grew larger and fewer, from 125 in 1820 to merely 54 at the eve of the Great War. Conversely, in the postwar era economic integration has been accompanied by political fragmentation, and the number of countries has risen to a record high of more than 190. At the same time, there has been a proliferation of international treaties and institutions aimed especially at fostering economic integration, such as the World Trade Organization and the European Union.

These trends are illustrated graphically in Figure 1, which shows the historical evolution of the number of sovereign states in the world, average exports as a share of GDP and the number of members of the GATT/WTO.[1] The sharp reversal in the link between economic and political integration presents an open puzzle and an obstacle for accounts of political geography based on economic efficiency. Does the evolution of the number and size of countries require explanations grounded in purely political or military considerations?

In this paper, we argue instead that the trends in Figure 1 are an efficient response to the falling cost of distance. Our starting point is that borders hamper trade and globalization makes borders costlier. Thus, political structure adapts by removing borders or reducing their cost. We show that this adaptation generates a non-monotonic evolution of efficient country size. In its early stages, the efficient political response to globalization consists of removing borders by creating large countries. In its later stages, the efficient political re-

[1] Data on the number of states was obtained from Butcher and Griffiths (2013). The trade share is from Maddison (2001). See Appendix A.1 for more details.

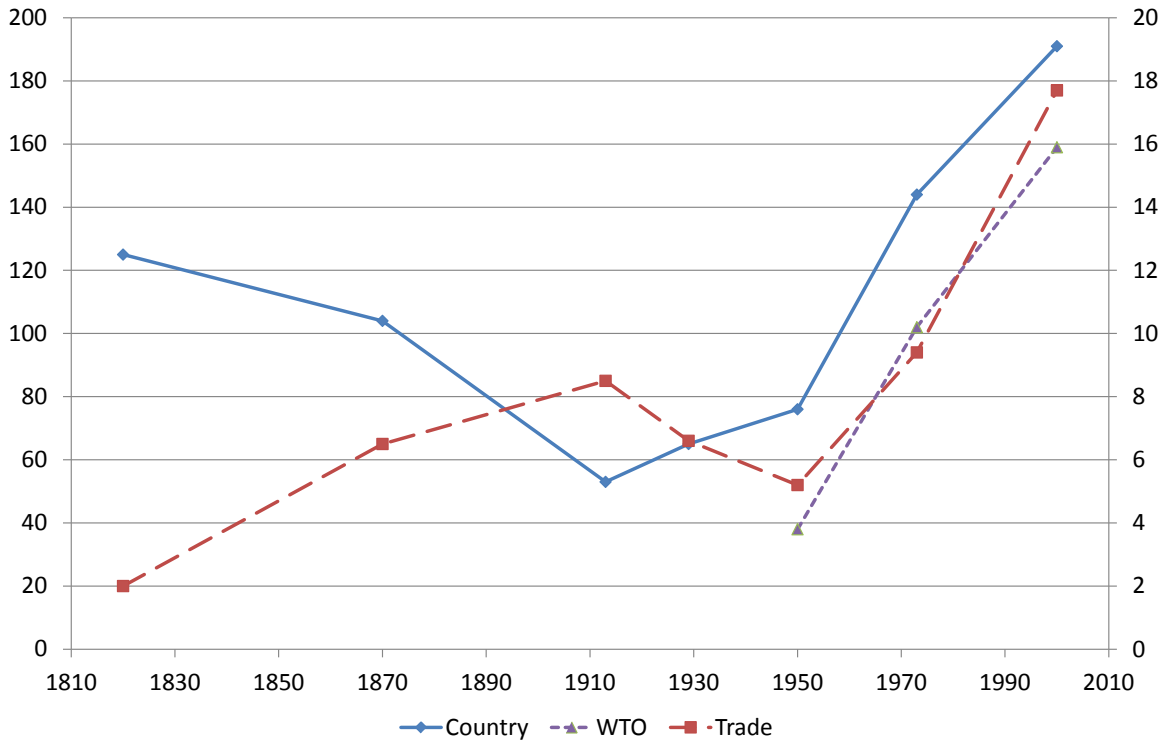

Figure 1: Economic Integration and Political Integration. The figure plots the trade share (right axis), the number of countries and of WTO members (left axis). See Appendix A.1 for details on data.

sponse to globalization is to remove the cost of borders by creating a new level of government, namely, international unions. This induces a reduction in the size of countries.

Naturally, our theory does not imply that political geography responds to economic efficiency alone. On the contrary, once we allow for inefficiencies our model sheds light on the patterns of interstate conflict too. When two-level governance is inefficient, we show it is appealing for the Great Powers to wage war to conquer markets. Empire-building enables them to trade with their colonies while imposing extractive institutions that deny them political autonomy. As efficient market size keeps growing, however, we find that this strategy reaches a breaking point. Conflict and imperialism are no longer advantageous for the Great Powers. They prefer gaining access to world markets through peaceful diplomacy, by building multilateral institutions that respect all members' political autonomy. These theoretical predictions are borne out by empirical evidence. During the first wave of globalization, increasing trade led to country expansion accompanied by conflict. Both links were broken during the second wave of globalization, which saw instead the peaceful rise of multi-level governance.

To derive our results, in section 2 we set up our model of a world with a continuum of basic geographical units or localities, each containing people that share common preferences. Goods can be transported at a negligible cost within localities, but at a positive cost across localities. Governments perform two tasks: (i) enforce contracts, protect property rights and enact economic regulations that help markets work; and (ii) provide public services such as education and welfare programs. We study how governments are organized geographically to perform these tasks and, in particular, how this organization changes as the cost of transporting goods across localities declines.

We make two standard assumptions about the effects of governments. The first is the presence of border effects. If two localities have different governments providing economic regulation, they can trade only a limited range of goods. The second assumption is preference heterogeneity over public services. If two localities have the same government, they receive the same basket of public services.[2]

Without costs of government, the optimal political structure would be a two-level governance structure. The first level would be a continuum of country governments, one for each locality, providing each of them with its preferred basket of public services. The second level would be a world government or international union that regulates markets and eliminates all border effects. Unfortunately, this political structure is too expensive.

We make two standard assumptions about the costs of government. The first is the presence of economies of scale. There are some costs of setting up and running a government that are fixed or independent of the number of localities sharing this government. The second assumption is the presence of economies of scope. Coordinating different levels of government is costly and, as a result, the two-level governance structure is more expensive than a single-level structure.

Economies of scale and scope affect the optimal political structure. Economies of scale make it desirable to have a discrete number of country governments rather than a continuum of them. Localities are willing to accept public services less than ideally tailored to their preferences in order to benefit from economies of scale. Economies of scope, if large enough, make it desirable to have single-level governance rather than a two-level structure. Localities

[2]We assume that local preferences differ with respect to public services but not economic regulation. Market-enabling economic regulation aims primarily at increasing efficiency, and this goal is widely shared by people with different preferences (Coase 1960; Posner [1973] 2014). Accordingly, government functions such as contract enforcement, monetary policy, or the policing of anti-competitive practices are often entrusted to apolitical technocrats. Public services, however, are a focus of political tension because people have different views on how children should be educated, on the proper size and scope of the welfare system, and so on. Such preferences vary systematically across localities reflecting their distinctive history and culture.

may also be willing to accept higher trade costs to take advantage of economies of scope.

Thus, the equilibrium political structure balances four classic forces: border effects, preference heterogeneity, economies of scale and economies of scope. In this context, we study how a reduction in transport costs affects this balance.

In Section 3, we assume that localities bargain efficiently in a world ruled by law and diplomacy. At early stages of globalization, the gains from trade are small and the benefit of creating an international union does not justify sacrificing economies of scope. Thus, a single-level governance structure is optimal. As globalization proceeds, the gains from trade grow, providing incentives to remove borders. The number of countries declines and the mismatch between each locality's ideal and actual provision of public services grows. Eventually, this mismatch is large enough to justify the move to a two-level governance structure. The world political structure shifts from a few large countries to many small countries within an international union. This two-level structure is more expensive, but it is nonetheless desirable because it facilitates trade and improves preference-matching in the provision of public services.

In Section 4 we allow a set of core localities to wage war and build empires. War is costly but it allows core localities to remove borders with the colonies while imposing their preferences on them. Thus, we find that there is an intermediate stage of globalization in which empires are formed. Eventually this stage ends, empires collapse, and international unions are created to promote trade and peace. Core localities choose to avoid the cost of war and to replace conquered colonies with free partners in an international union. The cause of imperial collapse at a late stage of globalization is exactly the same as the cause for the rise of empires at an early stage, namely, the desire to reap gains from trade in the most cost-efficient manner.

In Section 5, we use historical data to test the two most distinctive implications of our theory. The first is that increases in trade should predict territorial expansions before international unions are created, but not afterwards. The second is that territorial changes should be conflictual before international unions are created, but should be peaceful afterwards. Using data from the Cross-National Time-Series Archive from 1870 to 2010, we find indeed that increases in trade and proxies for military power predict subsequent territorial expansion before World War II, when international unions did not play a major role, but not after it. Moreover, we find that the incidence of conflict in border changes declines dramatically after World War II.

Section 6 concludes by using our theory to interpret a variety of historical experiences.

Related Literature The motivating facts presented in Figure 1 were first noted by Kahler and Lake (2004) and Lake and O'Mahony (2004). They highlight the puzzling reversal in the link between economic and political integration. They stress that it poses an obstacle for efficiency-based explanations of political structure, which have so far failed to account for it, as well as for the emergence of supranational institutions during the second but not the first wave of globalization.

Several economic theories can help explain why economic integration has been accompanied by political fragmentation since World War II (Bolton and Roland 1996, 1997; Alesina, Spolaore and Wacziarg 2000, 2005; Casella 2001; Casella and Feinstein 2002). All of them, however, predict a monotonic effect of globalization on political structure, and thus fail to explain why the first wave of globalization was accompanied by a decline in the number of countries. Nor can they explain the creation of international unions. Closest to our own work, Alesina, Spolaore and Wacziarg (2000, 2005) add border effects to Alesina and Spolaore's (1997) seminal theory of country formation based on the trade-off between preference heterogeneity and economies of scale.[3] They explain the increase in the number of countries during the second wave of globalization by interpreting globalization as an exogenous weakening of the border effect. As borders become less costly, efficient political structure reacts by creating more borders.[4]

Our model is the first to account for the non-monotonic impact of globalization on political structure, and to explain the appearance of multi-level governance during the second wave of globalization. The reason is twofold. First, we recognize that economies of scope are limited.[5] As a consequence, a broader set of political structures can be efficient, and we can explain the shift from a single-level to a two-level governance structure. Second, we consider a more primitive technological driver of globalization: the gradual decline of transportation costs. This change in focus enables us to show that globalization makes borders more costly rather than less. This creates incentives to remove borders rather than to create them. In our theory, the weakening of the border effect only occurs endogenously as political structure adapts to new trade opportunities. Both our innovations are crucial: without either,

[3] Desmet et al. (2012) calibrate this trade-off for European countries.

[4] Bolton and Roland (1996, 1997) focus on income distribution and find that heterogeneous countries may break up if their barriers to external trade decline. Casella (2001) and Casella and Feinstein (2002) study how preferences for public goods can endogenously become more heterogeneous as market size expands and enables greater specialization.

[5] Alesina and Spolaore (2003) offer an insightful discussion of economies of scope. Specifically, Chapter 2 analyzes an arbitrary system of overlapping jurisdictions, while Chapter 9 presents a system of overlapping jurisdictions constrained to form a pyramidal hierarchy.

theories based on economic efficiency would be unable to explain why economic and political integration went together until World War II, and then took different directions.

Our work is also related to the literature on trade and war. The idea that trade promotes peace was formalized by Polachek (1980). It is based on the premise that conflict harms trade and hence trade openness raises the opportunity cost of war (Alesina and Spolaore 2003; Rohner, Thoenig and Zilibotti 2013). The opposite idea that trade generates military conflict is instead expressed in neo-Marxist theories of imperialism. Trade can also make countries dependent on others and therefore vulnerable (Bonfatti and O'Rourke 2017).[6] Our paper suggests that these seemingly antithetical views capture two different stages of the same model. We provide a unified explanation for why territorial changes are more with military conflict in the first wave of globalization than in the second, an empirical pattern noted by Lake and O'Mahony (2006). Consistent with our result that international unions remove the incentive to wage war, Martin, Mayer and Thoenig (2012) find evidence that regional trade agreements promote peaceful relations.

The effect of war on country formation has been studied by Alesina and Spolaore (2003, 2005), Griffiths (2014), Gennaioli and Voth (2015) and Alesina, Reich and Riboni (2017), among others. These papers show that changes in military technology can explain country size, investment in state capacity and the provision of public goods. However, they find a monotonic effect of military technology on political structure. Our theory also recognizes conflict as one of the determinants of country formation and of the provision of public goods. In our model, waging war is one reason why countries grow large. We show, however, that changes in military technology alone are not a persuasive driving force for the reversal in the link between economic and political integration. Instead, we find that falling transportation costs can explain endogenously the switch from a world of aggression in which countries grow large to one of diplomacy in which countries become smaller.

Finally, our work is related more broadly to the economic analysis of federalism and of the geographic structure of government. Our model embeds the key trade-off that lies at the heart of the classic theory of fiscal federalism (Oates 1972). Centralization reaps economies of scale and benefits from policy coordination, but it imposes a uniform policy on localities with different preferences.[7] Models of political centralization and decentralization have

[6] Martin, Mayer and Thoenig (2008) show that while bilateral trade lowers the probability of conflict, multilateral trade openness decreases dependence on any given country and hence the cost of a bilateral conflict.

[7] Political-economy frictions micro-found these countervailing forces (Lockwood 2002; Besley and Coate 2003; Harstad 2007; Boffa, Piolatto and Ponzetto 2016). Multiple local governments fail to coordinate

been applied most often to the architecture of government at the sub-national level (Lockwood 2006; Treisman 2007). However, the same insights apply to the study of international unions (Hooghe and Marks 2003; Alesina, Angeloni and Etro 2005; Alesina, Angeloni and Schuknecht 2005; Ruta 2005). Prior research in this field has overwhelmingly focused on the optimal size and composition of an exogenously given number of government tiers. Surprisingly, the literature has devoted much less attention to the choice between a single-level and a multi-level governance structure, which our analysis focuses on.

2 A Model of Political Structure

In this section, we develop a stylized model of the world that contains the basic ingredients of our theory: geography, markets and preferences. The model mixes these ingredients imposing a high degree of symmetry. This allows us to derive our basic results on the effects of globalization on political structure quickly and intuitively.

The concept of locality is a key primitive in our theory. We model the world as a set of places within which there are neither geographical nor cultural distances, and we label them localities. Thus, localities consist of a group of people sharing common preferences and inhabiting a particular territory. This approach, which is common in the literature, simplifies the study of how peoples with different preferences interact and organize themselves into political entities. But it is silent about how these different preferences arose in the first place and how they evolve over time. It also abstracts from domestic conflict.

The concept of globalization is another important primitive in our theory. Geographical distances introduce trade costs across localities. In particular, we use the usual assumption of iceberg trade costs across localities. We interpret globalization as exogenous technological change that gradually removes these trade costs.

2.1 Basic Setup

We consider a world with a continuum of atomistic localities, $l \in [0, 1]$. Each locality contains a positive measure of identical individuals. W_l denotes the welfare of the representative individual of locality l. For short, we refer to this individual as "locality l." Then, the

efficiently even if they can bargain with one another. A single central government fails to match policies to local preferences even if it can differentiate policy across regions—in fact endogenous policy differentiation may prove more harmful than uniformity.

welfare of locality l is:

$$W_l = W_l^M + W_l^G, \tag{1}$$

where W_l^M is the utility derived from the consumption of market goods, and W_l^G is the utility derived from public services.

Governments provide public services and regulate markets, so government activity affects both welfare components. A political structure for this world consists of two partitions of the set of localities $[0,1]$ into governments: a public-service partition P with typical element $P_n \in P$; and an economic-regulation partition R with typical element $R_n \in R$.

If $P = R$, we say that the world has a single-level governance structure, and we refer to the common elements of P and R as country governments or countries. Each of these countries provides both public services and market regulation to its constituent localities.

If $P \neq R$, we say that the world has a two-level governance structure. It will presently become clear that governments have a pyramidal hierarchy: if the partitions P and R do not coincide, the finer partition P is always a refinement of the coarser one R. Hence, we refer to the (smaller) elements of P as country governments or countries, and the (larger) elements of R as international unions or unions. Countries provide public services to their constituent localities, while unions regulate the markets of their constituent countries.

We develop a model of the partitions P and R, that is, a model of how localities organize themselves into countries and how countries organize themselves into unions. We start from assumptions about preferences, technology and the costs of government, and we determine how welfare W_l depends on political structure (P, R).

2.1.1 Markets and Trade

There is a continuum of industries, $i \in [0,1]$. Let $c_l(i)$ be the consumption of goods of industry i by locality l. The utility function takes the following form:

$$W_l^M = \int_0^1 \ln c_l(i)\, di. \tag{2}$$

The production of final consumption goods requires industry-specific differentiated input varieties, $m \in [0,1]$. Define $c_l(m,i)$ as the amount of inputs of the variety m for industry i used by locality l in the production of final goods. Then:

$$c_l(i) = \exp\left\{\int_0^1 \ln c_l(m,i)\, dm\right\}. \tag{3}$$

These consumption preferences and production technology are symmetric across and within industries, and we specify convenient Cobb-Douglas functional forms. As a result, each locality spends an identical fraction of its income on each variety of each industry.

To introduce gains from specialization and trade, we adopt a simple symmetric version of the Ricardian model. Each locality is endowed with one unit of labor in each industry. This unit can produce one unit of the variety with the same index as the locality ($m = l$); or $e^{-\eta}$ units of any other variety ($m \neq l$). Since $\eta > 0$, each locality has a technological advantage in its "own" variety. The parameter η measures the extent to which technologies differ across localities and, therefore, the potential gains from specialization and trade.

There are technological barriers to trade. We assume uniform iceberg transportation costs across localities so that only a fraction $e^{-\tau} < 1$ of the goods shipped from l to $m \neq l$ arrives to destination. To focus on the most interesting case in which trade costs are not prohibitive and to ensure positive gains from trade, we assume that $\eta > \tau > 0$. Our measure of globalization is the wedge $\gamma \equiv \eta - \tau$, which captures the potential gains from trade and increases as improvements in transportation technology reduce physical trade costs τ. Globalization can thus range from $\gamma = 0$ when trade costs are prohibitive ($\tau = \eta$) to a maximum of $\gamma = \eta$ when trade costs are nil ($\tau = 0$).

Policy-induced barriers to trade or border effects arise when different governments regulate markets. In particular, we assume that exchanging goods in a fraction $\beta \in (0, 1)$ of industries requires legal enforcement of contracts. In these industries, varieties cannot be traded between localities that have different governments regulating their markets, i.e., that belong to different elements R_n and $R_{n'}$. The reason is that foreigners correctly anticipate that domestic courts will discriminate against them ex post. In the remaining set of industries, contracts are self-enforcing and thus varieties can be traded without restrictions. This formulation captures a simple and yet realistic micro-foundation for the well-known finding that borders obstruct trade.[8]

A market equilibrium is a set of prices and quantities such that individuals maximize utility and markets clear. Appendix A.2 shows that there exists a unique market equilibrium. Traded industries specialize in each locality's input variety, export essentially all of their production and import the remaining input varieties. Thus, consumption in traded industries is $c_l(m, i) = e^{-\tau}$. Nontraded industries are forced to produce locally all input varieties. Thus,

[8] This microfoundation is consistent with Broner and Ventura (2011). There are other microfoundations, though. For instance, tariffs and non-tariff barriers are also policies that discriminate against foreigners and limit the range of goods that can be traded.

consumption in nontraded industries is $c_l(m,i) = e^{-\eta}$. This implies the following utility from consuming market goods:

$$W_l^M = -\eta + \gamma\left(1 - \beta + \beta\int_0^1 I_{l=m}^R dm\right), \tag{4}$$

where $I_{l=m}^R$ is an indicator variable which takes value 1 if localities l and m belong to the same R_n, and zero otherwise. Equation (4) shows the impact of border effects. A decline in transportation costs raises the gains from trade γ in every industry. However, border effects prevent a locality from reaping the gains from trade in a mass β of industries that require contract enforcement. As a consequence, the value of removing each border effect is proportional to $\beta\gamma$, where γ measures the potential gains from trade in any single industry and β the mass of industries subject to border effects.

2.1.2 Governments

Public services consist of a basket of differentiated varieties, $x \in [0,1]$. The basket provided to locality l is characterized by a density function $g_l(x)$ defined over these varieties, with $g_l(x) \geq 0$ and $\int_0^1 g_l(x)\,dx = 1$. The utility derived from these public services is:

$$W_l^G = \int_0^1 \delta_l(x)\, u(g_l(x))\, dx - K, \tag{5}$$

where $\delta_l(x) \geq 0$, $u(g_l(x)) = -1/g_l(x)$ and K is a cost function to de defined shortly. We refer to the first and second terms of Equation (5) as the benefits and costs of public services respectively.

We now introduce three assumptions about governments. The first assumption is about preference heterogeneity. Each locality has a different ideal variety of public services. We define and order the basic varieties such that the ideal one for locality l is $x = l$. We assume that $\delta_l(x) = \delta$ if $x = l$; and $\delta_l(x) = 0$ otherwise.

The second assumption is that there are economies of scale in the provision of public services. Building and maintaining a government reduces the value or utility of public services by a total or fixed amount $\phi > 0$, and this cost is equally shared among the constituent localities.

Our third and final assumption is that there are economies of scope across government functions. Membership of a union reduces the value or utility of public services by an

amount $\kappa > 0$. This captures the costs of oversight and coordination between a country and the union.

These three assumptions imply the following utility from public services:

$$W_l^G = -\frac{\delta}{g_l(l)} - \frac{\phi}{\int_0^1 I_{l=m}^P dm} - \kappa I_l^U, \tag{6}$$

where $\delta > 0$; $I_{l=m}^P$ is an indicator variable that takes value 1 if if localities l and m belong to the same P_n, and zero otherwise; and I_l^U is an indicator variable that takes value 1 if locality l is a member of a union $R_n \neq P_n$, and zero otherwise. The first term in Equation (6) means that the value of public services for locality l depends on the amount of its ideal variety that is provided. The second term in Equation (6) means that each locality's share of the fixed cost of government declines with the size of the country. The parameter ϕ measures the magnitude of these economies of scale. The third term of Equation (6) means that being a member of a union is costly. The parameter κ measures the magnitude of these economies of scope.

To complete the model we need to make assumptions on how localities interact. We consider law and diplomacy in Section 3, and war and conquest in Section 4. In both cases, the world's political structure is determined by the interplay of the forces that follow from our assumptions. Although these assumptions are standard in the literature, we provide some additional discussion of them in Appendix A.3.

3 Law and Diplomacy

Efficient bargaining among localities delivers Pareto efficient outcomes and constitutes a natural benchmark to study. In this case, the equilibrium political structure is obtained by solving the following maximization problem:

$$(P, R) = \arg\max \int_0^1 \omega_l W_l dl, \tag{7}$$

where $\{\omega_l\}_{l\in[0,1]}$ is a set of Pareto weights such that $\int_0^1 \omega_l dl = 1$. Given the symmetry of this world, it seems reasonable to focus on the case in which the bargaining process treats all localities in the same way: $\omega_l = 1$ for all $l \in [0, 1]$. Sometimes this political structure is referred to as the utilitarian welfare optimum since it maximizes average world welfare. We view it as the description of a world in which all localities have the right to choose their own

political structure. This is a world ruled by law and diplomacy.

An implication of the maximization problem (7) is that each country P_n provides a uniform bundle that contains equal amounts of the ideal varieties of its constituent localities.[9] That is, locality l receives the following bundle of public services:

$$g_l(x) = \begin{cases} \dfrac{1}{\int_0^1 I^P_{l=m} dm} & \text{if } I^P_{l=x} = 1 \\ 0 & \text{if } I^P_{l=x} = 0. \end{cases} \tag{8}$$

Thus, we can re-write Equation (6) as follows:

$$W^G_l = -\delta \int_0^1 I^P_{l=m} dm - \frac{\phi}{\int_0^1 I^P_{l=m} dm} - \kappa I^U_l. \tag{9}$$

The first term means that the value of public services for locality l declines with the size of its country. As more localities join the country, the public services provided are farther away from the ideal of each member locality. The parameter δ measures the importance of this preference mismatch.

Combining Equations (1), (4), and (9), we obtain:

$$W_l = -\eta + \gamma \left(1 - \beta + \beta \int_0^1 I^R_{l=m} dm\right) - \delta \int_0^1 I^P_{l=m} dm - \frac{\phi}{\int_0^1 I^P_{l=m} dm} - \kappa I^U_l. \tag{10}$$

Equation (10) shows how political structure determines welfare and reveals the key trade-off that underlies our theory. A desirable political structure should facilitate trade, accommodate preference heterogeneity and take advantage of economies of scale and scope. But these goals cannot be achieved simultaneously and something must give.[10]

3.1 Equilibrium Political Structure

Two preliminary results simplify the analysis of the maximization problem (7). The first is that P and R contain equal-sized elements. Let S and U be the sizes of each element $P_n \in P$

[9] This is welfare-maximizing since localities have convex preferences.

[10] Equation (10) includes all the features of Oates's (1972) classic Decentralization Theorem: in the absence of cost savings from the centralized provision of public services ($\phi = 0$) and of interjurisdictional externalities ($\beta\gamma = 0$), welfare is at least as high if each locality can choose its own public services than if any uniform bundle is imposed across all of them.

and $R_n \in R$ respectively.[11] The second result is that, as anticipated, P is a refinement of R. If it is ever worth paying the costs of having a two-level governance structure, this is because localities desire a lower-level government that provides public services adapted to their specific preferences, and a higher-level government that reduces border effects and facilitates trade. Thus, we can write W_l as a function of S and U as follows:

$$W_l = W^F(S,U) = -\eta + \gamma(1-\beta+\beta U) - \delta S - \frac{\phi}{S} - \kappa I^U, \tag{11}$$

where I^U is an indicator variable that takes value 1 if $S \neq U$, and zero otherwise.

Equation (11) implies that the equilibrium political structure features either $P = R$ or $P \neq R = \{[0,1]\}$.[12] In the first case, the world is organized in a single-level governance structure with a set of countries that provide public services and regulate markets. In the second case, the world is organized in a two-level governance structure with countries providing public services and a world union regulating markets.

We now find the equilibrium political structure in three steps. First, we compute the welfare $W^F(S_1^*, S_1^*)$ generated by the single-level governance structure, where S_1^* is the optimal country size without a world union. This political structure takes full advantage of economies of scope, and country size trades off preference heterogeneity against both economies of scale and facilitating trade:

$$S_1^* = \sqrt{\frac{\phi}{\delta - \beta\gamma}}. \tag{12}$$

The size of countries in the absence of unions is increasing with economies of scale (ϕ) and the importance of trade ($\beta\gamma$), and it is decreasing with preference heterogeneity (δ).[13]

Second, we compute the welfare $W^F(S_2^*, 1)$ generated by a world with a union, where S_2^* is the optimal country size with a world union. This political structure gives up economies of scope in order to remove border effects and facilitate trade. Country size trades off preference

[11] Throughout, we disregard the constraint that the number of countries and unions, $1/S$ and $1/U$, must be a natural number. Aside from this constraint, all localities prefer the same optimal country size so that any equilibrium is symmetric. In Appendix A.4 we introduce this integer constraint and show that the equilibrium political strcuture remains symmetric and qualitatively analogous to the tractable approximation we use in the main text.

[12] We know that, if $P \neq R$, there is only one world economic union because the marginal cost of adding members is constant and the marginal benefit is growing with the size of the union. Thus, having many small unions is not optimal.

[13] Equation (12) assumes that $\delta > \phi + \beta\eta$, so that there is enough preference heterogeneity to ensure that countries are always smaller than the whole world.

heterogeneity and economies of scale:

$$S_2^* = \sqrt{\frac{\phi}{\delta}}. \tag{13}$$

The size of countries with a world union is increasing with economies of scale (ϕ) and it is decreasing with preference heterogeneity (δ). Country size is always smaller with a world union than without it. The reason is that the union removes one of the incentives for country size, namely, facilitating trade.

The third step is to determine the equilibrium political structure. If $W^F(S_1^*, S_1^*) > W^F(S_2^*, 1)$, the world is partitioned into countries of size S_1^*. If $W^F(S_1^*, S_1^*) < W^F(S_2^*, 1)$, the world is partitioned into countries of size S_2^* that belong to a world union. Naturally, in the knife-edge case in which $W^F(S_1^*, S_1^*) = W^F(S_2^*, 1)$, both solutions are equilibrium political structures. A little algebra shows that the world union is preferred if and only if:

$$\kappa + 2\sqrt{\phi\delta} < \beta\gamma + 2\sqrt{\phi(\delta - \beta\gamma)}. \tag{14}$$

That is, the world union is preferred for high values of β, γ and δ; and low values of ϕ and κ. A world union is more useful if the border effect and the gains from trade are large and there is substantial preference heterogeneity. A world union is less useful if economies of scale and scope are sizable.

3.2 Globalization and Political Structure

With these results at hand, we can now return to Figure 1 in the Introduction and ask again: Why did the first wave of globalization reduce the number of countries but not generate unions? Why did the second wave of globalization increase the number of countries and lead to the creation of unions? To answer these questions, we interpret globalization as a process by which γ grows from 0 to η, and we study how political structure changes as this process unfolds.

Figure 2 shows how equilibrium political structure depends on the two parameters that measure economies of scope and globalization, κ and γ. For a given κ, the world chooses a single-level political structure if γ is low. If κ is not too large, as in the dashed line, the world political structure shifts from single-level to two-level governance as globalization crosses a

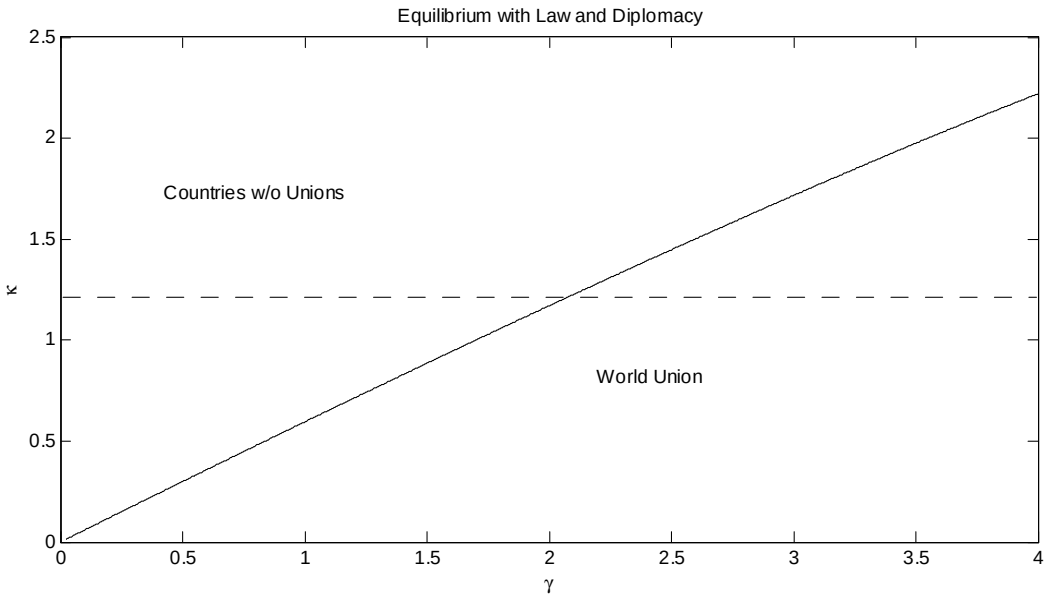

Figure 2: Globalization and Political Structure. The figure shows how equilibrium political structure depends on economies of scope (κ) and globalization (γ).

threshold value γ_U defined as follows:

$$\beta\gamma_U + 2\sqrt{\phi\left(\delta - \beta\gamma_U\right)} = \kappa + 2\sqrt{\phi\delta}. \tag{15}$$

If economies of scope are nil, the smallest gain from trade leads to the formation of a world union ($\kappa = 0$ implies $\gamma_U = 0$). If economies of scope are prohibitive, the world union is never an equilibrium ($\kappa > \beta\eta + 2\sqrt{\phi\left(\delta - \beta\eta\right)} - 2\sqrt{\phi\delta}$ implies $\gamma_U > \eta$). The comparative statics of this threshold follow directly from our analysis of Equation (14). The larger the border effect (β) and preference heterogeneity (δ), the smaller γ_U. The larger economies of scale (ϕ) and scope (κ), the larger γ_U.

Figure 3 shows how political structure changes with globalization by plotting the equilibrium size of countries and unions as a function of γ. At low levels of globalization ($\gamma < \gamma_U$), it is too expensive to create a world union, and increases in γ lead to an increase in country size. The cost of reaping additional gains from trade is a growing preference mismatch. Eventually, the preference mismatch has grown so large that it becomes cost-effective to create a world union. At high levels of globalization ($\gamma > \gamma_U$), the cost of reaping additional gains from trade is the loss of economies of scope. The creation of a world union allows countries to revert to a smaller size and reduce the preference mismatch. Further increases in γ have no effect on political structure.

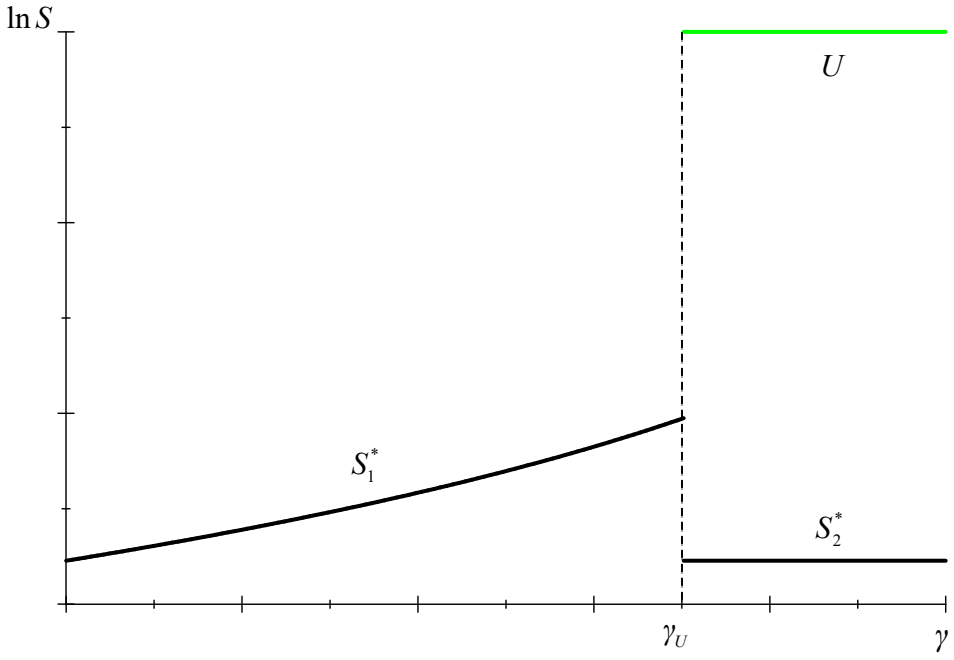

Figure 3: Globalization, Countries and Unions. The figure shows how the world political structure changes with globalization (τ). The black line is the size of each country, the green line is the world union.

4 War and Conquest

We explore next how war and conquest affect the relationship between globalization and political structure. To do this, we assume that the world is divided into core and periphery. The core contains a measure π of localities with a superior military technology that can be used to conquer other localities and form empires. The periphery contains the remaining localities that do not have this military technology. We assign low indices to core localities: $C = [0, \pi]$. We keep all assumptions regarding preferences, technology and the costs of government. Thus, the model of the previous section applies as $\pi \to 0$.

Empires are an alternative form of government that provides public services and regulates the markets. An empire contains a metropolis and its colonies. The metropolis consists of core localities that unite to conquer periphery localities that then become colonies. Empires

provide the ideal public services of their metropolises:

$$g_l(x) = \begin{cases} \dfrac{1}{\int_0^1 I^M_{l=m} dm} & \text{if } I^M_{l=m} = 1 \\ 0 & \text{if } I^M_{l=m} = 0, \end{cases} \tag{16}$$

where $I^M_{l=m}$ is an indicator variable that takes value 1 if localities l and m belong to the metropolis and zero otherwise. Empires provide none of the ideal public services of their colonies.

To build an empire of size E, core localities must wage war to conquer the colonies. War is successful if and only if the size of the metropolis M is large enough relative to that of the colonies:

$$M \geq \mu E, \tag{17}$$

where $\mu \in (0, 1)$. This assumption captures the well-known argument that country size is important for military success, and thus one of the reasons countries grow large is to prepare for war. We assume throughout that $\pi < \mu$. This means that the combined size of all empires is always smaller than the world, and empires do not need to fight each other for colonies.

From the perspective of the metropolis, the upside of building an empire is that it facilitates trade and generates economies of scale with minimal preference mismatch. The downside is that waging war and holding the empire together reduces the utility that the metropolis derives from public services by an amount $\omega > 0$. This cost captures the diversion of government resources from providing public services in the metropolis to waging colonial wars. Thus, the welfare of a member of the metropolis is:

$$W_l = -\eta + \gamma \left(1 - \beta + \beta \int_0^1 I^E_{l=m} dm\right) - \delta \int_0^1 I^M_{l=m} dm - \frac{\phi}{\int_0^1 I^E_{l=m} dm} - \omega, \tag{18}$$

where $I^E_{l=m}$ is an indicator variable that takes value 1 if localities l and m belong to the same empire and zero otherwise.

From the perspective of the conquered colonies, the gains from trade are dwarfed by the costs of an imperial government that generates an unbounded preference mismatch. This division of the surplus captures in our model the extractive nature of imperialism: the metropolis enjoys the gains from trade, while the conquered localities suffer under an exploitative colonial administration.

4.1 Equilibrium Political Structure

With war and conquest, equilibrium political structure need no longer be globally efficient. Formally, equilibrium political structure is now determined in two stages:

1. Core localities choose cooperatively whether to wage war and build empires.[14] Localities in the periphery may become their colonies or remain free.

2. Localities that do not belong to an empire choose their political structure through efficient bargaining, as in our baseline model.

The world's equilibrium political structure now consists of a set of empires that have a combined size $1-F$; plus two partitions (P, R) of the free world which itself has a combined size F. We solve for this equilibrium political structure in two steps. First, we determine the political structure of the free world (P, R) for a given size F. Second, we determine the number and size of empires and therefore the size of the free world F.

4.1.1 The Free World

The analysis of the free world is essentially the same as in the previous section. The only difference is that now the combined size of the free world is F rather than 1. Efficient bargaining ensures that free localities choose the optimal political structure. Equation (11) still applies and, as a result, there are two cases to consider: $S = U$ and $S < U = F$. The optimal country sizes in these cases are still given by Equations (12) and (13), respectively.[15]

The union of the free world is preferred now if and only if:

$$\kappa + 2\sqrt{\phi\delta} \leq \beta F\gamma + 2\sqrt{\phi\left(\delta - \beta\gamma\right)}. \tag{19}$$

Condition (19) generalizes Condition (14) for the case of a free world of size F. The main difference is that empires reduce the size of the free world, and this reduces the welfare associated with a union of free localities. This union still costs κ to each member. But it is now less efficient at removing border effects, βF instead of β.

[14] Historically, great powers have in fact cooperated and agreed on explicit partitions of the world, from the Treaty of Tordesillas in 1494 to the Berlin Conference in 1884.

[15] We now assume that $\delta > \phi F^{-2} + \beta\eta$ to ensure that countries are always smaller than the free world.

4.1.2 Empires

Core localities must first decide whether to wage war to build an empire or to forego war and enter the free world. We start the analysis with two observations. First, the equilibrium political structure features symmetric empires. Let E denote the size of each of them. The second observation is that constraint (17) is always binding. From the perspective of the metropolis, there is no reason not to add additional colonies. This lowers the cost of government and facilitates trade without creating any preference mismatch (for the metropolis) in the provision of public services.

These two observations imply that the welfare of a core locality that builds an empire is:

$$W_l = W^E(E) = -\eta + \gamma(1 - \beta + \beta E) - \delta\mu E - \frac{\phi}{E} - \omega. \tag{20}$$

This welfare does not depend on what other core localities do. This follows from our assumption that $\pi < \mu$, which ensures that core localities can find their desired measure of colonies without having to fight each other. The size of the empire trades off preference heterogeneity against both economies of scale and facilitating trade:

$$E^* = \sqrt{\frac{\phi}{\delta\mu - \beta\gamma}}, \tag{21}$$

where E^* is the optimal empire size (for the the core localities).[16]

Comparing Equation (21) to Equations (12) and (13), we see immediately that empires are larger than peaceful countries. The reason is that the metropolis does not internalize the cost of the preference mismatch imposed on the colonies: hence, $\delta\mu$ appears instead of δ in the denominator. The equilibrium size of empires is increasing with economies of scale (ϕ) and the importance of trade ($\beta\gamma$), and it is decreasing with preference heterogeneity (δ). These comparative statics are the same as for countries. We now have the additional result that the smaller the size of the metropolis relative to that of the colonies (μ), the larger the empire.

When are empires formed? If core localities wage war and build empires, their welfare is $W^E(E^*)$. If core localities instead agree to refrain from waging war and choose to form countries and unions by efficient bargaining, their welfare is $\max\left\{W^F(S_1^*, S_1^*), W^F(S_2^*, 1)\right\}$.

[16] Equation (21) assumes that $\delta\mu > \phi(\mu/\pi)^2 + \beta\eta$, so that there is enough preference heterogeneity to ensure that the metropolis is smaller than the whole core and a fortiori the empire is smaller than the whole world.

If $W^E(E^*) < \max\left\{W^F(S_1^*, S_1^*), W^F(S_2^*, 1)\right\}$, there are no empires, diplomacy rules and the size of the free world is $F = 1$. If instead $W^E(E^*) > \max\left\{W^F(S_1^*, S_1^*), W^F(S_2^*, 1)\right\}$, there are $\pi/(\mu E^*)$ empires of size E^*, and the size of the free world is reduced to $F = 1 - \pi/\mu$.[17] Some algebra shows that empires are built if:

$$\omega + \beta\gamma + 2\sqrt{\phi(\delta\mu - \beta\gamma)} \leq \min\left\{\beta\gamma + 2\sqrt{\phi(\delta - \beta\gamma)}, \kappa + 2\sqrt{\phi\delta}\right\}. \tag{22}$$

Unsurprisingly, empires are built in equilibrium if the military technology is good enough, i.e., the cost of waging war (ω) is low and the metropolis (μ) small relative to the empire.

To sum up, Condition (22) determines whether empires are built or not, and thus the size of the free world. Then, Condition (19) determines whether the free world is organized in a single-level governance structure or a two-level one.

4.2 Globalization and Political Structure

Let us now return to the question of how globalization affects political structure. Figure 4 shows again how equilibrium political structure depends on the two parameters that measure economies of scope and globalization, i.e. κ and γ. There is an age of empires if and only if the following condition holds:

$$\omega < 2\sqrt{\phi}\left(\sqrt{\delta - \beta\gamma_U} - \sqrt{\delta\mu - \beta\gamma_U}\right). \tag{23}$$

Otherwise, empires are never built and the world is ruled by law and diplomacy.

The dashed line in Figure 4 shows a scenario in which initially the whole world is free and there is no union. As globalization reaches a first threshold value $\gamma_L > 0$, empires are formed. Eventually, as globalization reaches a a second threshold value $\gamma_H < \eta$, empires are abandoned and a world union is formed. This evolution is not generic, though. A necessary and sufficient condition for $\gamma_L > 0$ is that:

$$\omega + 2\sqrt{\phi\delta\mu} > 2\sqrt{\phi\delta}. \tag{24}$$

[17] We see here the role played by the assumption that core localities cooperate. If core localities choose empires noncooperatively, there might be equilibria in which empires are formed when $\max\left\{W^F(S_1^*, S_1^*), W^F(S_2^*, 1)\right\} > W^E(E^*) > \max\left\{W^F(S_1^*, S_1^*), W^F(S_2^*, 1 - \pi/\mu)\right\}$. If core localities expect other core localities to build empires, their best response is to build an empire themselves. Once this happens, there is no incentive to deviate. This equilibrium is a coordination failure since it lowers the welfare of all the localities in the world. One can also construct equilibria with mixed strategies in which some core localities build empires and others do not. We ignore theses complications to simplify our analysis.

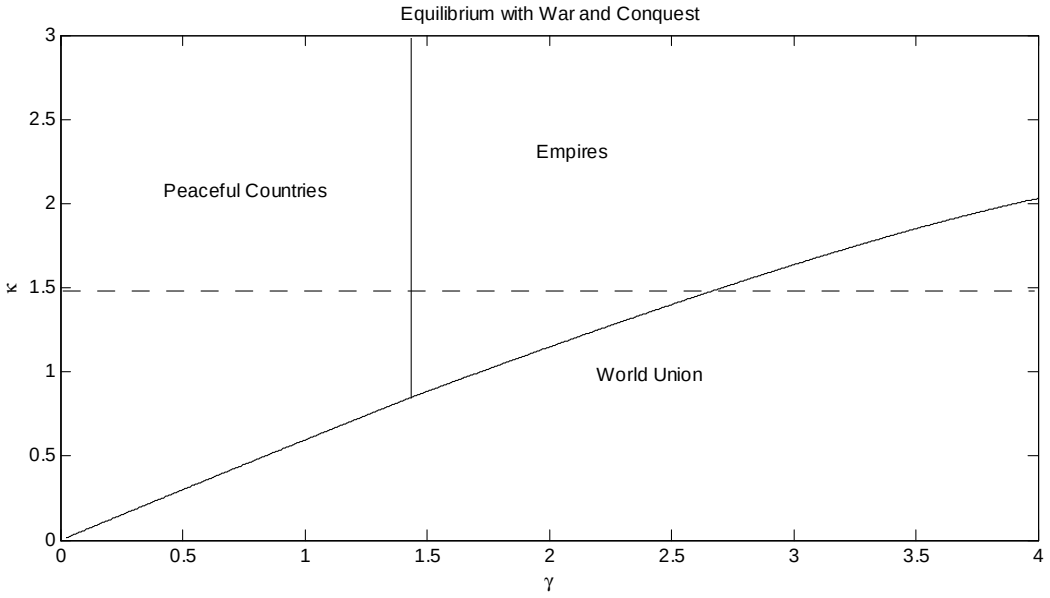

Figure 4: Globalization, Conflict and Political Structure. The figure shows how equilibrium political structure depends on economies of scope (κ) and globalization (γ).

If this condition holds, core localities do not conquer colonies just to share the fixed cost of government. Thus, core localities choose to forego warfare in autarky. It is only after the gains from trade have grown enough that empires become cost-effective from the point of view of the metropolis.

Also, a necessary and sufficient condition for $\gamma_H < \eta$ is that:

$$\omega + 2\sqrt{\phi\left(\delta\mu - \beta\eta\right)} > \kappa - \beta\eta + 2\sqrt{\phi\delta}. \tag{25}$$

If this condition holds, there is a level of globalization at which the preference mismatch has grown large enough within the metropolis to justify a move to the two-level governance structure. Paradoxically, the cause of imperial collapse at a late stage of globalization is exactly the same as the cause for the rise of empires at an early stage. Namely, the desire to remove border effects and reap the gains from trade. It is just that, at some point, it becomes more cost-efficient to replace conquered colonies with free partners in a union.[18]

[18]To simplify the analysis, we have ruled out by assumption the possibility that empires form economic unions. Yet, under mild assumptions, this can be obtained as a result. For example, suppose that core localities can impose their preferred political structure, including economic unions, onto the rest of the world. Then, under the condition $\omega > 2\sqrt{\phi\delta}(1-\sqrt{\mu})$, it is easy to show that core localities would strictly prefer diplomacy than war both in autarky and when the world union is in place. Hence, economic unions will be peaceful.

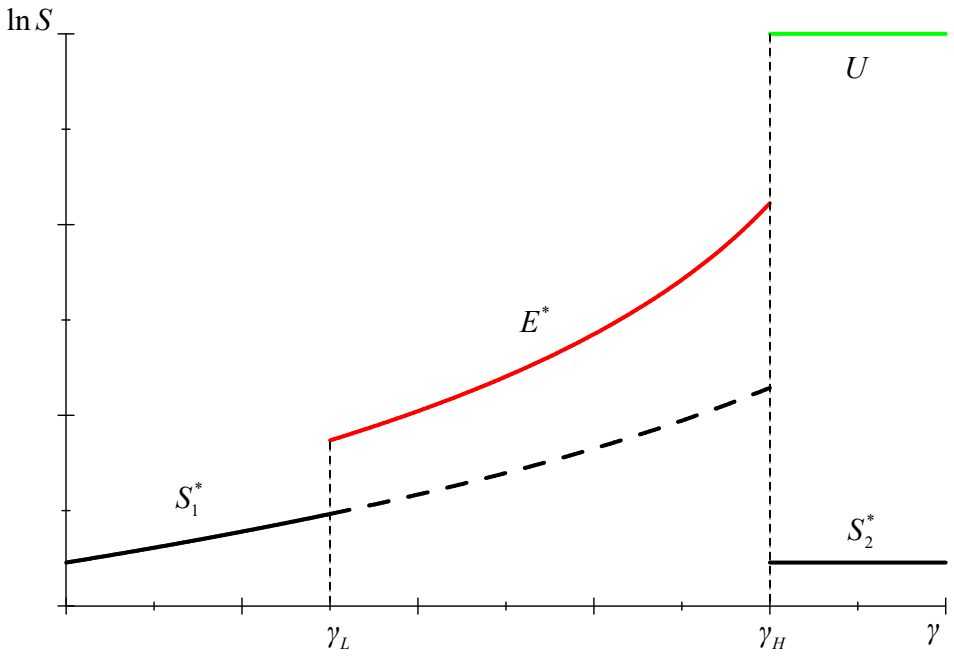

Figure 5: Countries, Empires and Unions. The figure shows how the world political structure changes with globalization (γ). The black line is the size of peaceful countries, the red line is the size of empires, the green line is the world union.

Figure 5 depicts this three-stage evolution of the world political structure by plotting the equilibrium size of empires, free countries, and unions. At low levels of globalization ($\gamma < \gamma_L$), the world contains only free countries. There are no empires or unions. As globalization proceeds, the size of countries grows. When globalization crosses the first threshold ($\gamma_L < \gamma < \gamma_H$), core localities prefer to build empires. Empires are larger than countries and keep growing as globalization proceeds. Eventually, globalization crosses the second threshold ($\gamma > \gamma_H$). Empires collapse and countries revert to a smaller size. A world union is created. After this, there are no further changes in political structure.[19]

Up to this point, we have focused exclusively on the effects of globalization (γ). But our model has other interesting parameters that capture preference heterogeneity (δ), the costs

[19]Figure 5 depicts the case in which, during the age of empires, the free world always adopts a single-level governance structure and the shift to the two-level governance structure coincides with the collapse of empires. This need not be the case. If the size of the free world is large enough, an economic union of free countries co-exists with empires.

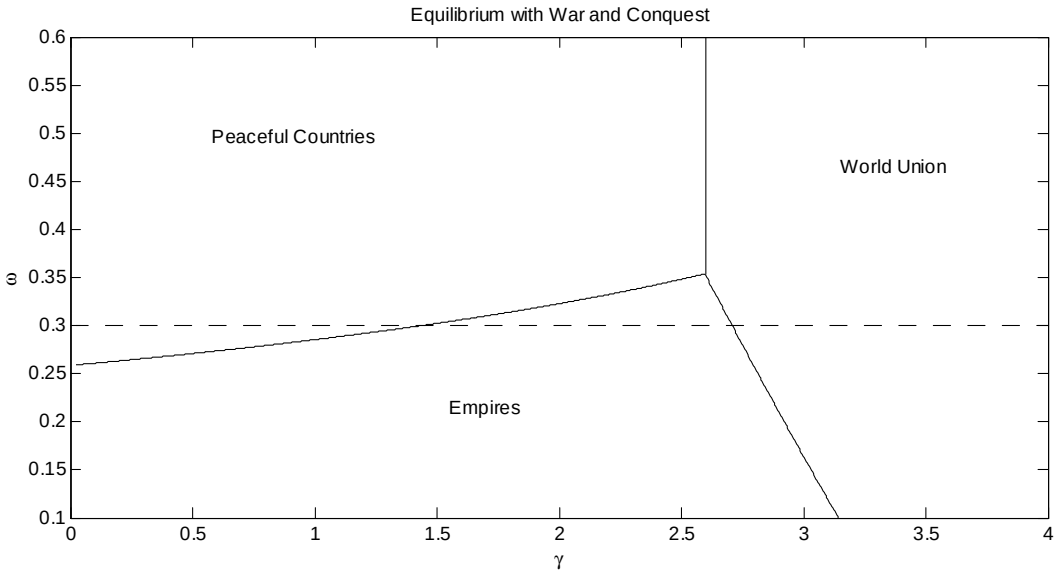

Figure 6: Trade, War and Political Structure. The figure shows how equilibrium political structure depends on the cost of war (ω) and globalization (γ).

of government (ϕ, κ, β) and military technology (ω, μ, π). Could changes in any of these parameters generate the three-stage evolution in political structure shown in Figure 5? It is not difficult to show that there is no change in any single parameter that would do so. For example, Figure 6 shows how equilibrium political structure depends on the cost of war and globalization, i.e. ω and γ. Changes in the cost of war, i.e. vertical movements for given γ, can explain the rise of empires or the switch to peaceful countries or unions. But they cannot explain a three-stage evolution, from countries to empires and from empires to unions.

4.3 Gradualism and Regionalism

War and conquest do not overturn our main result that globalization generates a shift in the world political structure from single-level governance to a two-level structure. Moreover, they explain why the creation and expansion of empires causes conflict and aggression during the first wave of globalization. Our model also explains why the formation of unions during the second wave of globalization promotes peace.

Perhaps the most unrealistic aspects of our model are its lack of regionalism and its lack of gradualism. If core localities join a union, it is a world union. If a world union is created, country size is reduced at once to its autarky level. However, these predictions are not robust to a slightly richer treatment of geography. Assume that, in addition to having

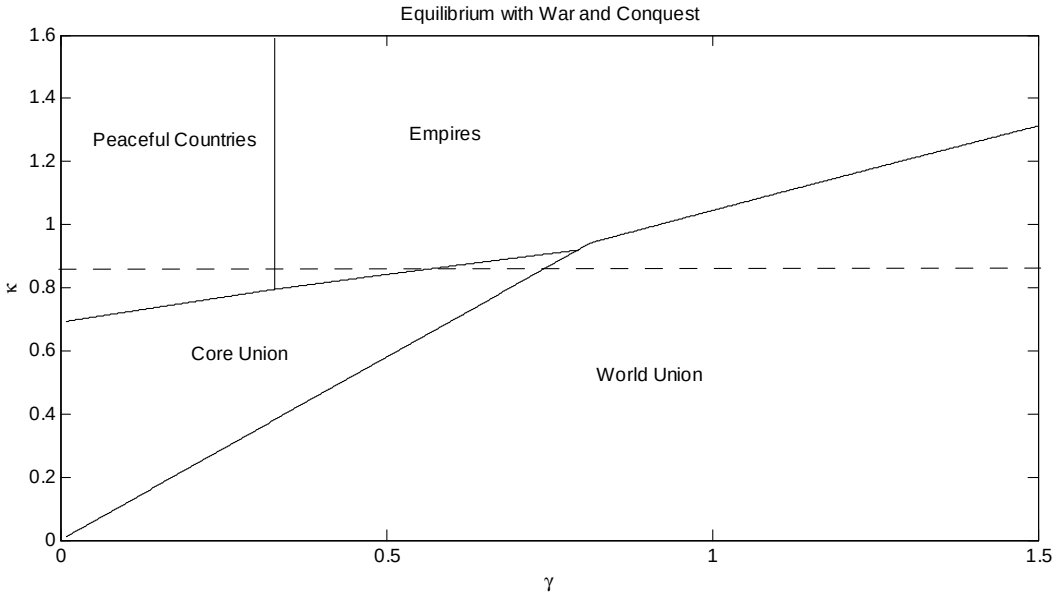

Figure 7: Gradualism and Regionalism. The figure shows how equilibrium political structure depends on economies of scope (κ) and globalization (γ).

a superior military technology, core localities are near each other. Transportation costs for core-core trade are $\tau - \rho$ (with $\rho > 0$), so that the gains from core-core trade are now $\gamma + \rho$. Transportation costs for core-periphery and periphery-periphery trade remain τ, so the gains from these types of trade are still γ. The rest of our assumptions remain the same.[20]

With this additional assumption, a mixed political structure with two-level governance in the core and single-level governance in the periphery becomes possible. Figure 7, which assumes a large enough value for ρ, shows a scenario in which globalization triggers a four-fold evolution. The first stage is peaceful single-level governance. The second is the creation of colonial empires through which the core conquers and rules distant localities in the periphery. As globalization proceeds further, empires become overstretched and collapse, and a peaceful union replaces them. The novelty of this third stage is that, unlike in Figure 4, this is now

[20] Appendix A.5 provides a full analysis of this model. Here we just state the new results it generates. Technically, the key implication of this new assumption is that Equation (4) must be replaced by:

$$W_l^M = -\eta + \gamma\left(1 - \beta + \beta\int_0^1 I_{l=m}^R dm\right) + \rho I_{l\in C}\left[(1-\beta)\,\pi + \beta\int_0^{\pi} I_{l=m}^R dm\right],$$

where $I_{l\in C}$ is an indicator variable that equals 1 if $l \in C$ and zero otherwise. Border effects are larger for core-core trade, $\beta(\gamma+\rho)$, than for either core-periphery or periphery-periphery trade, $\beta\gamma$. Other things equal, core localities prefer sharing economic regulation with other core localities. Periphery localities, instead, are indifferent about which localities they share regulation with and care only about their total number.

a core union. Only as globalization progresses further a fourth and final stage is reached in which the whole world forms a union.

The model can be further extended in fruitful ways. One extension is a world with two, three or N peripheries that are located progressively farther away. In this world, there is a union that starts at the core and grows outwardly with globalization. When the first periphery joins the union, the size of its countries declines. When the second periphery joins the union, the size of its countries also declines, and so on. The union gradually advances outward and it keeps breaking up countries. The end point is the same as in our baseline model, but the world approaches it gradually.

Another extension is a world with two or more core-periphery structures, which we can think of as continents or regions. In this world, within each region there is one union that advances outward, breaking up countries. Across regions, however, there is no union initially. Eventually, globalization may go so far that a world union becomes cost-effective, and the regional unions merge. The world approaches the same end point again, but it now approaches it both gradually and regionally.

5 Historical Evidence on Trade, Conflict and Country Size

Our theory provides a new perspective on the connection between trade, conflict, country size and the emergence of international unions. Our motivation for developing this theory was to improve our understanding of global trends, as presented in Figure 1. But the mechanisms that connect trade and country size over time should also be at work when we compare the trajectories of different countries. Thus, we now examine cross-country historical data in search for evidence on their empirical relevance. The results, though far from being conclusive, suggest that the mechanisms our theory highlights are indeed empirically relevant.

5.1 Trade and Country Size

The most fundamental implication of our theory is that changes in country size should be driven, at least in part, by changes in the volume of trade. In particular, the most novel result of our model is that increases in the volume of trade should predict increases in country size until the formation of international unions, and possibly decreases in country

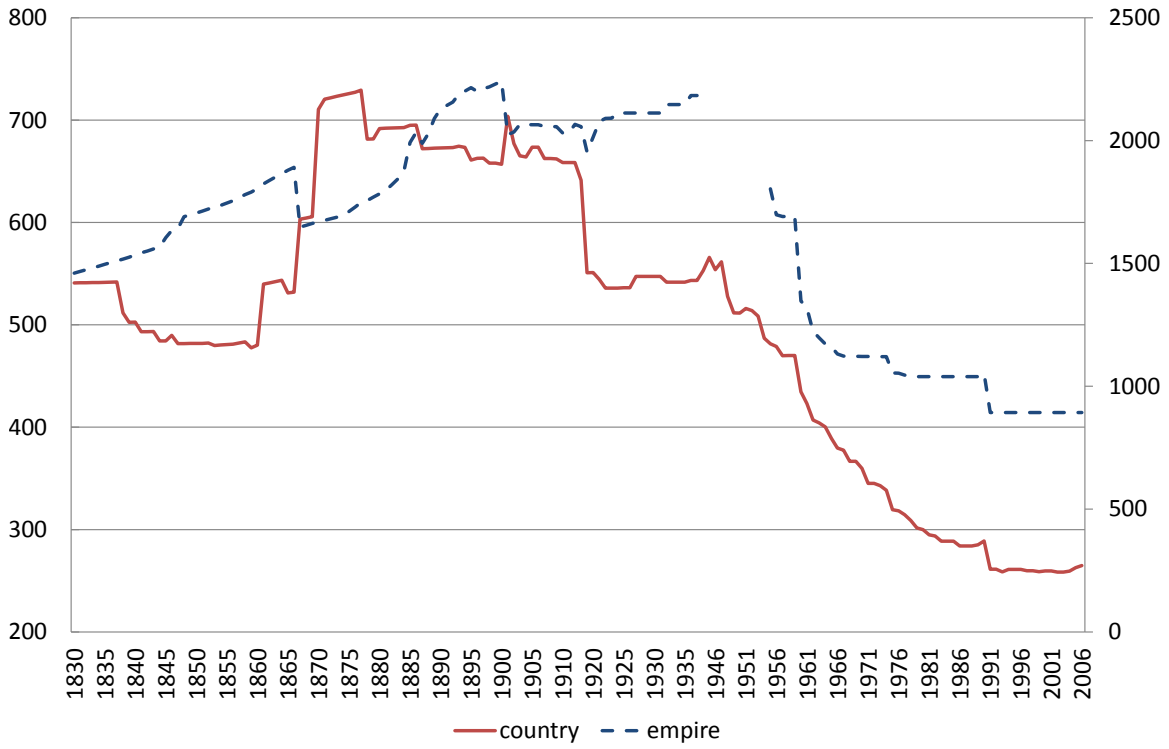

Figure 8: The Size of Countries and Empires. The figure plots the average size of countries (left axis) and of empires (right axis) in thousand squared miles. See Appendix A.1 for details on data.

size afterwards.[21] This reversal in the connection between changes in trade and changes country size is what differentiates our theory from others. Since international unions start to play an important role after World War II, we expect to find a positive correlation between trade and country size before World War II, which however becomes smaller, or even turns negative, after 1945.

To test this prediction, we use the Cross-National Time-Series (CNTS) Data Archive. This dataset contains land area and information on the volume of trade, measured as the sum of imports and export per capita, for an unbalanced panel of countries with observations from 1870 to 2010.[22] Using this dataset, Figure 8 shows the average land area of "internationally recognized" countries and of thirteen major empires since 1830. The size of both countries and empires increases until the appearance of international unions and falls thereafter. Empires, besides being larger than countries, start reducing their size somewhat

[21] In our model, trade eventually triggers the formation of international unions, and at that point country size falls discontinuously. Yet, given that countries join unions gradually, more trade may still raise the size of countries that are not part of unions.

[22] Before 1870, trade data is missing for the majority of countries.

Table 1: Descriptive Statistics

Period	Countries	Land Area	Expansion	Trade Growth	
	Number	Mean	Mean	Mean	Std. Dev.
1870-1880	43	911,867	0.233	0.280	0.421
1880-1890	45	974,517	0.244	0.271	0.430
1890-1900	46	985,006	0.217	0.084	0.303
1900-1910	46	922,332	0.087	0.403	0.346
1910-1920	47	959,006	0.234	1.410	1.262
1920-1930	55	865,532	0.164	0.466	1.041
1950-1960	73	591,716	0.041	0.880	0.971
1960-1970	102	460,871	0.010	1.043	1.268
1970-1980	127	377,772	0.024	4.613	4.504
1980-1990	149	274,399	0.027	0.276	0.668
1990-2000	158	257,461	0.006	1.851	14.732
2000-2010	169	294,461	0.012	1.698	1.327

Notes: Countries and Land Area are measured at the beginning of each decade. Land Area is expressed in thousand square miles. Expansion is a dummy taking value 1 if a country's land area expandend over the decade. Trade Growth is expressed in per capita terms.

later than countries. This picture is consistent with the data on the number of countries shown in Figure 1. Despite the different data sources, both figures tell a remarkably similar story.[23] During the nineteenth century there was a phase of political concentration in which countries and empires expanded their territories. But this trend reversed during the twentieth century, and especially after World War II.

Table 1 contains descriptive statistics for the main variables of interest. For each decade, it reports the number of countries with non-missing observations, their average land area, the share of countries that expanded their territories, the average change in the volume of trade and its standard deviation. Given that all years corresponding to the two world wars have no observations in the dataset, the decades around 1940 are missing. A quick look at Table 1 confirms the basic trends already discussed. Average land area increases in the first part of the sample, but this trend is clearly reversed after 1950. Accordingly, the share of countries expanding their territories falls dramatically after 1950. Average trade grows throughout

[23] Figure 8 is based on the "International System" (Singer and Small 1966), which includes countries with international recognition and reliable data. See Appendix A.1 for more details. This classification however underestimates the number of independent political entities in the nineteenth-century developing world. Figure 1 is instead based on Butcher and Griffiths (2013), who provide a more comprehensive record.

the entire period, but at different speeds both over the decades and across countries.

Given that land area changes slowly and discontinuously, we focus on decades and build a dummy variable that takes value 1 if a country or empire has grown in size over the previous ten years. We interpret this variable as the "probability" of a change in country size or territorial expansion. To study how this probability depends on the change in the volume of trade, we run a series of binomial regressions. To alleviate simultaneity, we compute the change in the volume of trade over the previous decade. Furthermore, to test if the correlation between trade and territorial expansion changes after World War II, we include an interaction term between changes in the volume of trade and a post-1950 dummy.

We use territorial expansions rather than territorial contractions because of data quality. Almost by definition, for any territorial expansion there are one or more corresponding territorial contractions.[24] But most territorial expansions occur in core countries, for which data are abundant and reliable. Instead, most territorial contractions occur in periphery countries, for which data are scarce and unreliable.[25]

The main results are reported in Table 2. Standard errors are corrected for clustering by country, so as to accommodate autocorrelated shocks at the country level. In columns 1 to 4, we start by excluding the interwar years and estimate a logit model. Initially, we only include as regressors the lagged change in the volume of trade and its interaction with the post-1950 dummy. Consistent with our theory, the coefficient for the lagged change in trade is positive, the interaction term is negative, and both are highly significant. Next, we also include the post-1950 dummy. The inclusion of the constant and the post-1950 dummy mean that the identifying variation is deviations from the global trends visible in Figures 1 and 8. Nevertheless, the two coefficients of interest remain significant. We then add sequentially the level of population and urbanization at the beginning of each decade. The first variable controls for the effect of size while the second is a proxy for economic development, and both are likely to be correlated with the military strength of a country.[26] Both coefficients are positive and significant, but their inclusion does not affect the trade variables.

In column 5 we maximize sample size by adding all available decades and a time dummy for the interwar period. Although the coefficient for trade falls in size, it remains significantly different from zero. In columns 6 we add time dummies. While time dummies control for

[24] Basically all of the relevant territories were already claimed by some country in 1870. Therefore, for each country that expands, there must be one or more countries that contract (or even disappear).

[25] We performed all the regressions that follow also for territorial contractions. Unsurprisingly given the low quality of the data, none of the variables was significative in any of the regressions.

[26] More direct measures, such as GDP, are not available for the entire period of analysis.

Table 2: Trade and Territorial Expansion. Logit and OLS regressions.

Dependent variable: Expansion dummy									
	Logit (1)	Logit (2)	Logit (3)	Logit (4)	Logit (5)	Logit (6)	Logit (7)	OLS (8)	OLS (9)
Trade Growth	2.283***	0.821*	0.818*	0.847*	0.362*	0.526**	0.760**	0.080*	0.076*
	(0.658)	(0.453)	(0.451)	(0.449)	(0.210)	(0.255)	(0.367)	(0.044)	(0.042)
Trade Growth × Post 1945	-2.512***	-0.821*	-0.821*	-0.862*	-0.379*	-0.858***	-1.794***	-0.080*	-0.077*
	(0.660)	(0.454)	(0.454)	(0.452)	(0.213)	(0.308)	(0.529)	(0.044)	(0.042)
Post 1945		-2.432***	-2.552***	-2.975***	-3.228***	-5.596***	-2.498	-0.590***	-0.396*
		(0.444)	(0.468)	(0.601)	(0.608)	(1.389)	(2.695)	(0.163)	(0.217)
Population			0.003***	0.003***	0.003***	0.004***	0.014	0.000***	-0.000
			(0.001)	(0.001)	(0.001)	(0.001)	(0.011)	(0.000)	(0.001)
Urbanization				0.002*	0.003**	0.004***	-0.006	0.000**	-0.000
				(0.001)	(0.001)	(0.001)	(0.006)	(0.000)	(0.000)
Interwar dummy					-0.300				
					(0.355)				
Constant	-2.882***	-1.533***	-1.599***	-1.693***	-1.595***	-0.098	-1.053	0.572***	0.402**
	(0.237)	(0.286)	(0.287)	(0.277)	(0.248)	(0.659)	(1.240)	(0.160)	(0.193)
Country FE	No	No	No	No	No	No	Yes	No	Yes
Time FE	No	No	No	No	No	Yes	Yes	Yes	Yes
Observations	733	733	733	713	799	799	264	799	799
R^2	0.088	0.200	0.217	0.230	0.230	0.280	0.369	0.181	0.345

Notes: All observations refer to 10-year periods. The dependent variable is a dummy taking value 1 if the country's land area expanded over the decade and 0 otherwise. Trade Growth is expressed in 10-year lags. Post 1945 is a dummy for decades after 1945. All other variables are measured at the beginning of each decade. In columns (1)-(4), the period from 1910 to 1950 is excluded. Pseudo-R^2 are reported for logit estimations. Standard errors, clustered by country, are in brackets. *, ** and *** denote significance at 10%, 5% and 1% respectively.

shocks affecting all countries, including to some extent globalization, they do not affect our main results. In fact, the coefficients for the trade variables become even more significant. In column 7 we add country fixed effects. This specification is exceedingly demanding: all countries with no changes in size are dropped and the coefficients are identified only from within-country deviations from country-specific trends. Nonetheless, trade and its interaction with the post-1950 dummy are the only two coefficients that remain highly significant.

Interpreting the magnitude of the coefficients in a logit model is difficult. Moreover, estimating a fixed-effect model for non-linear regressions can be problematic. For these reasons, and as an additional robustness check, in columns 8-9 we re-estimate the specifications in 6-7 using a linear probability model. The coefficients for trade are still significant. Their mag-

Table 3: Pre 1945 vs. Post 1945. Logit regressions.

Dependent variable: Expansion dummy								
	Pre 1945 (1)	Pre 1945 (2)	Pre 1945 (3)	Pre 1945 (4)	Post 1945 (5)	Post 1945 (6)	Post 1945 (7)	Post 1945 (8)
Trade Growth	0.431*	0.610**	0.474**	0.704***	-0.006	-0.213*	-0.066	-0.191*
	(0.231)	(0.262)	(0.232)	(0.258)	(0.020)	(0.124)	(0.113)	(0.104)
Population	0.018**	0.017**			0.003***	0.003***		
	(0.008)	(0.008)			(0.001)	(0.001)		
Urbanization	0.004**	0.006***	0.004**	0.006***	0.000	0.001	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Interwar dummy	-0.476		-0.605					
	(0.421)		(0.455)					
Size of Military			0.002	0.002*			0.000	0.000
			(0.001)	(0.001)			(0.000)	(0.000)
Constant	-2.010***	-0.622	-1.983***	-0.652	-4.177***	-4.476***	-4.014***	-4.648***
	(0.304)	(0.640)	(0.310)	(0.680)	(0.508)	(1.219)	(0.851)	(1.418)
Time FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	225	225	219	219	574	574	270	270
R^2	0.087	0.149	0.099	0.165	0.042	0.089	0.026	0.043

Notes: All observations refer to 10-year periods. The dependent variable is a dummy taking value 1 if the country's land area expanded over the decade and 0 otherwise. Trade Growth is expressed in 10-year lags. Post 1945 is a dummy for years after 1945. All other variables are measured at the beginning of each decade. In columns (1)-(4), the period from 1910 to 1950 is excluded. Pseudo-R^2 are reported for logit estimations. Standard errors, clustered by country, are in brackets. *, ** and *** denote significance at 10%, 5% and 1% respectively.

nitude implies that a doubling in the volume of trade during the previous decades increases the probability of a territorial expansion in a decade before 1950 by around 8%.

5.2 Conflict and Country Size

Another distinctive implication of our analysis is that territorial changes should be associated with conflict before international unions are created, but should be peaceful afterwards. Our model does not merely account for the switch from a pre-war age of expanding countries and conflict to a post-war era of shrinking countries and peace. The more novel element that differentiates our theory from others is that in the age of international unions even increases in country size should no longer reflect conflict and military might.

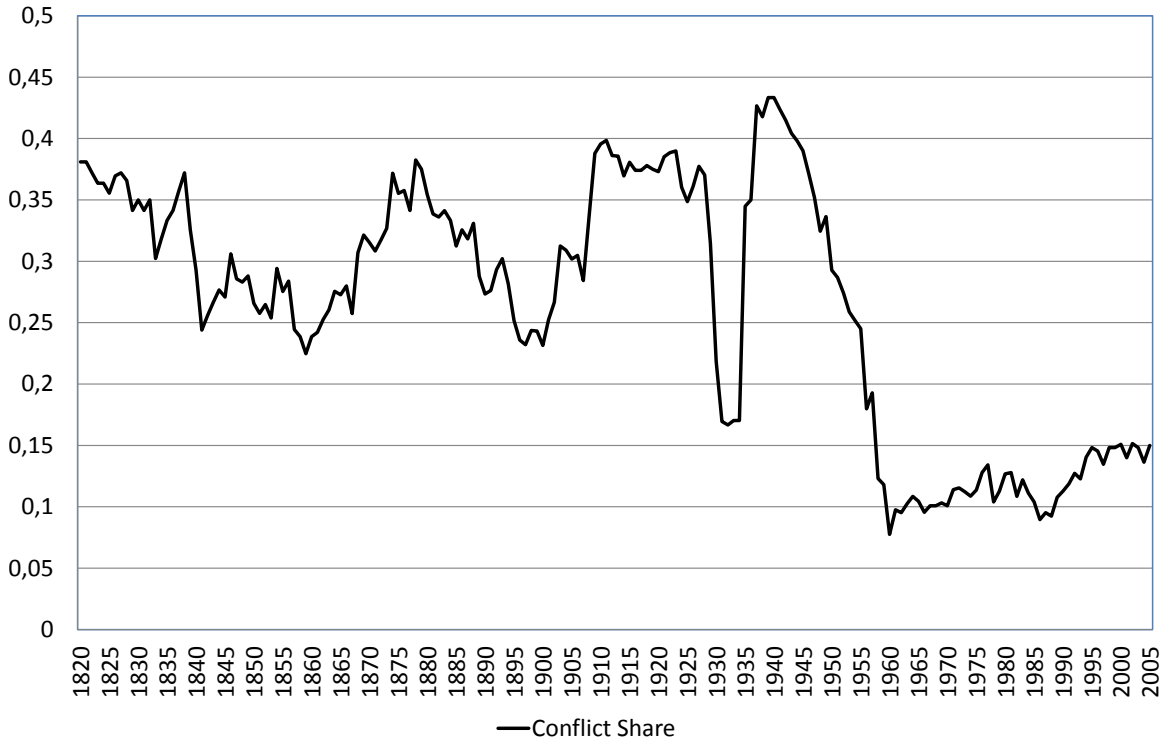

Figure 9: Border Changes and Conflict. The figure plots the share of changes of borders involving military conflict. Each point corresponds to an average over a +/-10 year window. Source: Correlates of War.

We start exploring the empirical support for this prediction by testing if our proxies for military strength, population and urbanization, have a different effect in the two sample periods. To this end, in Table 3 we split the sample and re-estimate the specifications with and without time effects (corresponding to columns 5 and 6 in Table 2) separately for the decades before 1940 (columns 1-2) and after 1950 (columns 5-6). To use a more direct proxy of military strength, in columns 3-4 and 7-8 we also replace population with the size of a country's armed forces.[27] Consistent with our theory, all these proxies for military strength have much larger and more significant coefficients in the pre-1940 period. The results in Table 3 also allow a direct comparison of the net effect of trade on territorial expansion in the two sub-periods. Not only we find that the growth in trade predicts subsequent territorial expansions in the per 1940 period, but also that the coefficient turns negative, and sometimes statistically significant, after 1950.

Finally, we complement this statistical analysis with more descriptive evidence. Figure 9

[27]This is the military variable with the highest coverage in the dataset. Its correlation with population precludes including both variables.

illustrates the incidence of violence in the historical process of redrawing political borders. It displays the share of territorial changes that involved military conflict among all territorial changes recorded over a period of ten years before and after each date, from the Correlates of War project. Consistent with our theory, the figure clearly shows that the end of World War II marks a structural break: before 1950, more than one third of all territorial disputes were decided by war, while after that date diplomacy prevailed in almost 90% of cases.

6 Narratives

We have shown that the expansion of trade opportunities can help explain two salient aspects of the evolution of political structure over the last couple of centuries: (i) the rise and subsequent fall in the size of countries observed during the nineteenth and twentieth century, and (ii) the seemingly contradictory trends towards more political integration across countries and more political fragmentation within countries in the second half of the twentieth century. We have also provided some historical evidence in support of this claim.

We conclude the paper, on a more speculative vein, by showing how our theory can help interpret historical events.

European History Since the late Middle Ages, European sovereign states on average grew in size until the end of the nineteenth century, when this trend was dramatically reversed. For example, Kitamura and Lagerlöf (2016) show that borders declined monotonically from 1500 to 1900, and then started to increase. Medieval Europe was fragmented into hundreds of small states at a time when trade was costly, insecure and limited to few commodities. The early modern period saw important changes in both the economic and the political organization of the continent. With the Commercial Revolution, trade began to flourish and the feudal system started to be replaced by a smaller number of countries of growing size. While in 1600 there were 112 sovereign states in Europe and the Near East, at the beginning of 1800 the number had fallen to 79.

The Industrial Revolution gave trade an even more prominent role and triggered major changes in socioeconomic conditions that ultimately made the rise of the nation state possible. Trade expansion was enabled by the introduction of canals, improved roads and railways. At the same time, the high degree of political fragmentation at the time of the Congress of Vienna (1815) was followed by the unification of Germany and Italy (1871) and the further consolidation of other nation states. The economic rationale of building large internal markets was especially evident in the case of German unification, which started

with the formation of a customs union (*Zollverein*). This process of political centralization culminated at the beginning of the twentieth century, when Europe was dominated by just 28 independent states.

Yet, the twentieth century marks a turning point. Europe entered a stage of political fragmentation, with the number of independent states growing to 58 in 2000. Simultaneously, the process of creating the European Union (EU) started. The core-periphery model shows that international unions start from countries that have closer economic ties and expands outward as globalization increases the value of trade with more remote locations as well. This prediction is consistent with the history of the EU. In 1957, Belgium, France, Italy, Luxembourg, the Netherlands and West Germany signed the Treaty of Rome establishing the European Economic Community (EEC) as a customs union. The EEC gradually expanded to include nearby countries: Denmark, Ireland and the United Kingdom in 1973, Greece in 1981, Portugal and Spain in 1986. In 1992, the Maastricht Treaty converted the EEC into the EU and, in 1995, Austria, Finland, and Sweden joined. In 2002, the Euro was launched and since then the EU has grown to include 28 member states in 2013. Over this period, the union has grown in size and scope.

In June 2016, however, the United Kingdom voted in a referendum to leave the union and there is a concern that other countries might follow in the future. While rising nationalism might have played a role, it is instructive to look at these events from the lens of our model. Three lessons can be learned. First, the value of union membership is proportional to the economic ties between countries. These are stronger for countries located in the core of continental Europe. For instance, while almost 80% of Belgium's total exports are delivered to other EU partners, the same figure is around 50% for the United Kingdom. Second, our model provides a rationale for trade-promoting unions. As more power is shifted to the union in other areas such as migration, tension may arise, especially in countries with a strong national identity. These two observations may explain why some UK politicians have advanced the idea of replacing the EU market with a Commonwealth free-trade zone. Third, the model suggests that the value of joining the union is proportional to its size. As a country exits, the economic foundations of the union become more fragile.

The Rise and Fall of Colonial Empires The pattern of an initial decline and subsequent increase in the number of countries is not confined to Europe only. For instance, the number of African countries fell from 36 in 1816 to 4 in 1914, to rise again to 51 in 2000. Similarly, in South-East Asia, these numbers changed from 37 to 4 and then 20 in the years 1816,

1914 and 2000, respectively.[28] However, to better interpret the political evolutions in these regions, it is important to bring conflict into the picture.

In our model of war and conquest, empires are built to extract trade surplus from the colonies and disappear when the union is formed to foster free markets. According to historians and in line with this view, one of the key driving forces behind colonial expansion was the desire to secure trade and access to scarce resources in an era of revived commerce, but when mercantilist practices where common. In fact, for much of the second millennium, states deployed force to create markets (Findlay and O'Rourke, 2007). Due to the scarcity of land and the desire to avoid powerful rivals, European great powers expanded by conquering territories overseas. The role of colonial powers in enforcing trade within the empire but not outside was very clear in the case of maritime commerce. On the one hand, large naval forces were built to control and protect trading routes; on the other hand, privateers were often authorized to capture merchant ships belonging to enemy nations. Despite some notable setbacks, colonialism continued to grow prior to World War I and finally collapsed after World War II.

The sharp decline of empires started after the creation of international agreements aimed at promoting economic cooperation. This is no coincidence. In the words of Rosecrance (1986) and Spruyt (2005), empires dissolved, often peacefully, because the gains through commerce displaced gains through territorial acquisition.[29] There is also evidence that international organizations played a direct role in the process of decolonization. For example, in 1960 the UN General Assembly voted the Declaration on the Granting of Independence to Colonial Countries and Peoples. Interestingly, our model of regional unions is consistent with the very different patterns of geographic expansions of empires and unions. As long as war is the dominant means of territorial expansion, the European Great Powers try to avoid conflict with each other by conquering far-away colonies and building global empires. Yet, once trade is enforced by peaceful international unions, countries seek economic integration with proximate partners, with whom the gains from trade are higher. The switch from global empires to regional unions is also consistent with the increased regionalization of world trade patterns observed in the data (e.g., Fouquin and Hugot 2016).

[28] The number of countries is taken from Butcher and Griffiths (2013).

[29] Bonfatti (2012) also attributes the fall of empires to the growing importance of trade between industrial countries relative to trade with colonies.

The United States Improvements in transportation technology and the desire to create a large internal market were important factors in its westward expansion. The abundance of land made it possible to create one of the largest countries in the world, without the need to seek far-away colonies. Despite its size, the United States avoided the phase of collapse and political fragmentation by choosing an institutional system with multiple levels of government. In this light, the experience of the United States follows the main pattern predicted by our theory: the creation and expansion of the federal government, which can be interpreted as a regional union, coincided with the fragmentation and loss of political power of individual states.

After the Declaration of Independence, the borders of the original thirteen states extended up to the Mississippi river, while the remaining land was occupied by French and Spanish colonies (later on part of Mexico) and by many tribes of native Americans living essentially in autarky. As the federal government acquired land and built roads westward, its territory was gradually fragmented into the fifty states. This process followed a common pattern. First, new land was annexed as large "territories;" subsequently this land was broken into new states. Federal expansion was followed not just by the creation of new states, but also by the break-up of existing ones. For example, Georgia, Massachusetts, North Carolina and Virginia all lost land to form new states. So did the former Republic of Texas (an independent country until 1846), which encompassed large parts of current Oklahoma, Kansas, Colorado, Wyoming and New Mexico. This pattern is consistent with our core-periphery model, in which the outward expansion of the core union breaks up countries in the periphery.

References

[1] Alesina, Alberto, Ignazio Angeloni and Federico Etro (2005). "International Unions," *American Economic Review* 95(3): 602–15.

[2] Alesina, Alberto, Ignazio Angeloni, and Ludger Schuknecht (2005). "What Does the European Union Do?" *Public Choice* 123(3): 275–319.

[3] Alesina, Alberto, Bryony Reich and Alessandro Riboni (2017). "Nation Building, Nationalism and Wars," NBER Working Paper No. 23435.

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

[5] —— (2003). *The Size of Nations*. Cambridge, MA: MIT Press.

[6] —— (2005). "War, Peace, and the Size of Countries." *Journal of Public Economics* 89(7): 1333–1354.

[7] Alesina, Alberto, Enrico Spolaore and Romain Wacziarg (2000). "Economic Integration and Political Disintegration," *American Economic Review* 90(5): 1276–1296.

[8] —— (2005). "Trade, Growth and the Size of Countries," in *Handbook of Economic Growth*, Vol. 1, Part B, edited by Philippe Aghion and Steven N. Durlauf, 1500–1542. Amsterdam: North Holland.

[9] Besley, Timothy, and Stephen Coate (2003). "Centralized versus Decentralized Provision of Local Public Goods: A Political Economy Approach," *Journal of Public Economic* 87(12): 2611–2637.

[10] Boffa, Federico, Amedeo Piolatto, and Giacomo A. M. Ponzetto (2016). "Political Centralization and Government Accountability," *Quarterly Journal of Economics* 131(1): 381–422.

[11] Bolton, Patrick, and Gerard Roland (1996). "Distributional Conflicts, Factor Mobility and Political Integration," *American Economic Review* 86(2): 99–104.

[12] —— (1997). "The Breakup of Nations: A Political Economy Analysis," *Quarterly Journal of Economics* 112(4): 1057–1090.

[13] Bonfatti, Roberto (2012). "Trade and the Pattern of European Imperialism, 1492–2000," Oxford University Department of Economics Discussion Paper No. 618.

[14] Bonfatti, Roberto and Kevin H. O'Rourke (2017). "Growth, Import Dependence and War," *Economic Journal*, forthcoming.

[15] Broner, Fernando and Jaume Ventura (2011). "Globalization and Risk Sharing," *Review of Economic Studies* 78(1): 49-82.

[16] Butcher, Charles, and Ryan Griffiths (2013). "Introducing the International System(s) Dataset (ISD), 1816–2011," *International Interactions* 39(5): 748–768.

[17] Casella, Alessandra (2001). "The Role of Market Size in the Formation of Jurisdictions," *Review of Economic Studies* 68(1): 83–108.

[18] Casella, Alessandra, and Feinstein, Jonathan S. (2002). "Public Goods in Trade: On the Formation of Markets and Jurisdictions," *International Economic Review* 43(2): 437–462.

[19] Coase, Ronald H. (1960). "The Problem of Social Cost," *Journal of Law and Economics* 3: 1–44.

[20] Desmet, Klaus, Michel Le Breton, Ignacio Ortuño-Ortín and Shlomo Weber (2011). "The Stability and Breakup of Nations: A Quantitative Analysis," *Journal of Economic Growth* 16(3): 183–213.

[21] Findlay, Ronald and Kevin H. O'Rourke (2007). *Power and Plenty: Trade, War and the World Economy in the Second Millennium*. Princeton, NJ: Princeton University Press.

[22] Fouquin, Michel and Jules Hugot (2016). "Back to the Future: Trade Costs and the Two Globalizations," CEPII Working Paper No. 2016-13.

[23] Gennaioli, Nicola and Joachim Voth (2015). "State Capacity and Military Conflict," *Review of Economic Studies* 82(4): 1409–1448.

[24] Griffiths, Ryan (2014). "Secession and the Invisible Hand of the International System," *Review of International Studies* 40(3): 559–581.

[25] Harstad, Bård (2007). "Harmonization and Side Payments in Political Cooperation," *American Economic Review* 97(3): 871–889.

[26] Hooghe, Liesbet, and Gary Marks (2003). "Unraveling the Central State, but How? Types of Multi-Level Governance," *American Political Science Review* 97(2): 233–243.

[27] Kahler, Miles, and David A. Lake (2004). "Governance in a Global Economy: Political Authority in Transition," *PS: Political Science & Politics* 37(3): 409–414.

[28] Kitamura, Shuhei, and Nils-Petter Lagerlöf (2016). "Geography and State Fragmentation," Working Paper, Institute for International Economic Studies.

[29] Lake, David A., and Angela O'Mahony (2004). "The Incredible Shrinking State: Explaining Change in the Territorial Size of Countries," *Journal of Conflict Resolution*, 48(5): 699–722.

[30] Lake, David A., and Angela O'Mahony (2006). "Territory and War: State Size and Patterns of Interstate Conflict," in *Territoriality and Conflict in an Era of Globalization*, edited by Miles Kahler and Barbara F. Walter, 133–55. Cambridge: Cambridge University Press.

[31] Lockwood, Benjamin (2002). "Distributive Politics and the Costs of Centralization," *Review of Economic Studies* 69(2): 313–337.

[32] Lockwood, Benjamin (2006). "The Political Economy of Decentralization." In *Handbook of Fiscal Federalism*, edited by Ehtisham Ahmad and Giorgio Brosio, 33–60. Cheltenham: Edward Elgar.

[33] Maddison, Angus (2001). *The World Economy: A Millennial Perspective*. Paris: OECD.

[34] Martin, Philippe, Thierry Mayer and Mathias Thoenig (2008). "Make Trade Not War?" *Review of Economic Studies* 75(3): 865–900.

[35] —— (2012). "The Geography of Conflicts and Regional Trade Agreements," *American Economic Journal: Macroeconomics* 4(4): 1–35.

[36] Oates, Wallace E. (1972). *Fiscal Federalism*. New York, NY: Harcourt Brace Jovanovich.

[37] Polachek, Solomon W. (1980). "Conflict and Trade," *Journal of Conflict Resolution* 24(1): 55–78.

[38] Posner, Richard A. (2014). *Economic Analysis of Law*, 9th edition [1st edition 1973]. New York, NY: Wolters Kluwer Law & Business.

[39] Rohner, Dominic, Mathias Thoenig and Fabrizio Zilibotti (2013). "War Signals: A Theory of Trade, Trust, and Conflict," *Review of Economic Studies*, 80(3): 1114–1147.

[40] Rosecrance, Richard (1986). *The Rise of the Trading State: Commerce and Conquest in the Modern World*. New York, NY: Basic Books.

[41] Ruta, Michele (2005). "Economic Theories of Political (Dis)integration," *Journal of Economic Surveys* 19(1): 1–21.

[42] Singer, David J., and Melvin Small (1966). "The Composition and Status Ordering of the International System: 1815–1940," *World Politics* 18(2): 236–282.

[43] Spruyt, Hendrik (2005). *Ending Empire: Contested Sovereignty and Territorial Partition.* Ithaca, NY: Cornell University Press.

[44] Treisman, Daniel (2007). *The Architecture of Government: Rethinking Political Decentralization.* Cambridge: Cambridge University Press.

A Appendix

A.1 Data Sources

The trade share reported in Figure 1 is merchandise exports as percent of GDP in 1990 prices, from Maddison (2001). Maddison provides trade data for selected countries in the years 1820, 1870, 1913, 1929, 1950, 1973 and 1998. To avoid compositional effects, we report the value of merchandise export as a share of GDP for the set of countries with data for all the years (Austria, Belgium, France, Italy, Spain, Switzerland, the United Kingdom and the United States). The trade share computed using the data for all available countries in every year is very similar to the one displayed in Figure 1.

The number of countries is reported for the same years. Data on the number of countries in the twentieth century is not very controversial. For the nineteenth century, however, some leading conventions grossly underestimate the number of countries. For example, the "International System," developed by Singer and Small (1966) and adopted in the Correlates of War project or in the Cross-National Time-Series Data Archive, only includes countries with international recognition. In particular, prior to 1920, the criteria to be recognized as an independent country were to have population greater than 500,000 and to have had diplomatic missions at or above the rank of chargé d'affaires with Britain and France. Clearly, this definition is too strict for our purposes, which require the identification of even relatively small political units living in economic and political autarky. We follow Butcher and Griffiths (2013), who recognize the problem and offer alternative criteria to identify the number of countries between 1816 and 2011.

The number of WTO members is from the WTO website.

The size of countries and empires displayed in Figure 8 is from the Cross-National Time-Series Data Archive (CNTS). It provides data on contiguous territorial area in thousand square miles for all countries existing in a given year according to the International State System. In a few instances, missing data have been imputed cross-checking major territorial changes from other sources (China and Persia before 1860). Area of empire is provided for a consistent sample of 13 countries: Austria (formerly Austria-Hungary), Belgium, France, Germany (formerly Prussia), Italy (formerly Sardinia), Japan, Netherlands, Portugal, Russia, Spain, Turkey (formerly the Ottoman Empire), United Kingdom, and United States. For these countries, empire area includes "overseas" territories (i.e., colonies). Data for the two World War periods, 1914-1918 and 1940-1945 (1938-1954 for Empires) are missing.

For the analysis in Section 5, the volume of trade is computed as the value of imports

plus exports per capita. Population is in millions. Urbanization is aggregate population in cities of 100,000 and over divided by total population. For empires, trade, population and urbanization refer to the metropolis only. Land area includes oversea territories.

The number of independent states in Europe and Near East reported in Section 5 is taken from Euratlas-Nüssli (http://www.euratlas.com).

A.2 Computing Equilibrium Consumption

Locality l maximizes the objective function:

$$W_l^M = \int_0^1 \int_0^1 \ln c_l(m,i)\, dmdi, \tag{A1}$$

subject to the budget constraint:

$$\int_0^1 \int_0^1 p_l(m,i)\left[c_l(m,i) - q_l(m,i)\right] dmdi \leq 0, \tag{A2}$$

where $q_l(m,i)$ and $p_l(m,i)$ are the production and price of input m of industry i in locality l. The productions $q_l(m,i)$ must be consistent with available technology as described in the text. Since individuals are atomistic, they take prices as given in their maximization problems.

We claim now that equilibrium prices are given as follows:

$$p_l(m,i) = \begin{cases} 1 & \text{if } l = m \\ e^{\eta} & \text{if } i \in [0,\beta] \text{ and } I_{l=m}^R = 0 \\ e^{\tau} & \text{if } i \in (\beta,1] \text{ or } I_{l=m}^R = 1 \text{ but } l \neq m. \end{cases} \tag{A3}$$

To prove this claim, normalize world income to unity ($Y = 1$). Note first that each locality has unit density of expenditure on each input in each industry. We next examine production. Consider first industries that require contract enforcement, $i \in [0,\beta]$. Locality l employs unit density of labor to produce each non-traded input m for which $I_{l=m}^R = 0$. Thus, output of each non-traded input has density $e^{-\eta}$, so the value of output has unit density given price e^{η}. The remaining mass $\int_0^1 I_{l=m}^R dm$ of industry-i labor is employed to produce an identical mass of output. Unit density of it is sold domestically at a unit price. The remainder is shipped in identical amounts to other localities with $I_{l=m}^R = 1$, each of which receives a density $e^{-\tau}$ of imports, hence import value of unit density given price e^{τ}. In industries that do not

require contract enforcement, $i \in (\beta, 1]$, the whole unit mass of industry-i labor is employed to produce the locality's own input variety, which is sold in identical amounts to all other localities in the world. Thus, the value of sales in each locality of each input in each industry has unit density, just like expenditure. This proves our claim.[30]

With these prices at hand, we can compute the equilibrium productions and consumptions described in the text.

A.3 Discussion of Assumptions

The starting point of our analysis is the idea that improvements in transportation technology are a major driver of globalization, which we model as a fall in trade costs between localities. This view is uncontroversial: there is overwhelming evidence that the secular rise in trade volumes was made possible by better transportation technologies. For instance, some of the major drivers of the first wave of globalization are the adoption of the steamship (Pascali 2017), the telegraph (Steinwender 2018) and the spread of railroads (Donaldson 2018). Likewise, the main drivers of the second wave of globalization include the use of containerization in ocean shipping (Levinson 2006), the development of jet aircraft engines (Hummels 2007) and more recently the ICT revolution. All these innovations promoted trade both between and within countries.

We also assume that borders obstruct trade. The large negative effect of political borders on trade volumes is well know at least since the work of McCallum (1995), who showed that, controlling for distance and income, trade between two Canadian provinces is 20 times larger than trade between a Canadian province and a U.S. state. While the exact magnitude of the border effect is still subject to debate, all existing studies coincide in finding large effects. For instance, in a recent survey of the voluminous empirical literature on gravity equations, Head and Mayer (2014) report that countries are typically found to trade 5 to 7 times more with themselves than with any other country.

There is equally strong evidence that sharing economic regulations and signing economic agreements promote trade and reduce the border effect. For instance, Head and Mayer (2014) also report that sharing a common currency or being part of a free trade area are associated on average with a doubling of the volume of trade; similarly, Helpman et al.

[30] It is straightforward to show that this equilibrium is unique. First, rule out variation in the prices of traded inputs since this would generate excess demand (supply) of cheap (expensive) varieties. Second, rule out that the relative prices of traded and nontraded varieties be above (below) τ/η since this would lead to an excess demand (supply) of nontraded inputs.

(2008) find that having a similar legal system increases the bilateral volume of trade by more than 60 percent. Using a simple model, Anderson and van Wincoop (2004) attempt a rough decomposition of the border effect. They argue that the compounded cost of borders is equivalent to an ad valorem tax of 44 percent, which can be broken down into an 8 percent of policy related barriers (including non-tariff barriers), 7 percent language barriers, 14 percent currency barriers, 6 percent information cost barriers, and 3 percent security barriers.

Our modeling assumption that technological barriers reduce trade along the intensive margin while policy-induced barriers affect the extensive margin is also grounded in empirical evidence. There is a wide consensus that transportation costs affect significantly the intensive margin of trade; on the other hand, Helpman et al. (2008) and Dutt et al. (2013) find that free-trade agreements and WTO membership predominantly affect the extensive margin.

Turning next to governments, our model follows the standard assumptions that underpin the literature on federalism and the architecture of government since Oates (1972).[31] Having separate local governments enables better preference matching, but sharing a common government enables beneficial policy coordination and reaps economies of scale. While these assumptions originated as simple observations of real-world patterns, models of political economy have provided them with rigorous micro-foundations. Majority rule makes centralization costly when localities have different preferences (Lockwood 2002; Besley and Coate 2003). Frictions in bargaining between political leaders (Harstad 2007) and in their agency relationship to their constituents (Boffa, Piolatto and Ponzetto 2016) explain both why multiple local government cannot fully coordinate their policies and why a single central government cannot fully tailor public services to local preferences. Accordingly, Strumpf and Oberholzer-Gee (2002) find empirically that U.S. states with more heterogeneous preferences are more likely to decentralize policy-making, while Lassen and Serritzlew (2011) show that municipal amalgamations cause citizens to experience a worse fit between what they want and what they get from local government. Evidence from sub-national governments also confirms the existence of economies of scale in government administration, albeit—consistent with our model—not in the provision of public goods more broadly (Reingewertz 2012; Blom-Hansen, Houlberg, Serritzlew 2014; Blom-Hansen et al. 2016; Blesse and Baskaran 2016).

Our assumption of economies of scope in government is equally classic (Musgrave 1971;

[31] The same forces that operate in our model at the national and supra-national level should also apply at the sub-national level. We could explore this issue further by introducing a third set of government functions that would justify a three-level governance structure. We conjecture that, in this setting, increasing globalization would explain simultaneously the creation of international unions and the trend towards greater federalism within countries.

Dahl and Tufte 1973; Alesina and Spolaore 2003). Marks and Hooghe (2004, p. 18) "emphasize the costs of decomposing authority" as a paramount concern in the analysis of multi-level governance, especially in the international arena with its prevalence of intersecting task-specific jurisdictions. Empirical evidence shows that multiplying administrative tiers reduces their efficiency, and is particularly associated with lower labor productivity and excess government employment (Le Galès and John 1997; Andrews and Boyne 2009). This cost is particularly pronounced for special-purpose governments in charge of a single task (Berry 2009). These efficiency losses reflect both the costs of administrative duplication and economies of scope in political accountability. Boffa, Piolatto and Ponzetto (2016) show theoretically that dividing policy-making responsibilities across multiple levels of government increases overall rent extraction by government officials. Fan, Lin and Treisman (2009) report that across countries corruption increases with the number of administrative tiers: as they rise from two to six, the probability of a firm reporting that it is never expected to pay bribes falls by 32 percentage points.[32]

A.4 Dealing with Integer Constraints

A world without unions consists of $N^* \in \mathbb{N}$ countries. Country n consists of measure $S_n > 0$ of localities, such that $\sum_{n=1}^{N^*} S_n = 1$ and utilitarian social welfare is $W = \sum_{n=1}^{N^*} S_n W^F(S_n, S_n)$.

The welfare function

$$W^F(S_n, S_n) = -\eta + \gamma(1 - \beta + \beta S_n) - \delta S_n - \frac{\phi}{S_n} \tag{A4}$$

is concave in S_n and increasing at $S_n = 0$. Whenever $\delta > \phi + \beta\gamma$ it has a unique maximum at $S_n = S_1^*$.

Pareto efficiency then requires that either $S_n \leq S_1^*$ for all $n = 1, 2, ..., N$ or $S_n \geq S_1^*$ for all n. Otherwise some localities could leave a country with excessive size $S_n > S_1^*$ and join another with insufficient size $S_n < S_1^*$, raising the welfare of every locality in both countries.

Utilitarian welfare maximization requires all countries to have the same size. If there are two countries m and n such that $S_m > S_n > S_1^*$, then transferring the marginal locality from m to n not only raises its welfare, but it also raises the welfare of S_m localities by more than

[32] Admittedly, the evidence about the importance of economies of scope has been gathered mostly at the sub-national level. Casual observation suggests, however, that economies of scope also apply at the supranational level. A notorious example is the European Central Bank, which is aimed at reducing the border effect by eliminating currency barriers. Its creation does not seem to have reduced the size or costs of national central banks, but instead it seems to have just added to these costs.

it lowers the welfare of $S_n < S_m$ localities. Likewise if $S_m < S_n < S_1^*$.

Therefore, once integer constraints are taken into account, the utilitarian welfare optimum without unions is a partition of the world into a number

$$N_1^* = \arg\max_{N \in \mathbb{N}} \left\{ -\eta + \gamma \left(1 - \beta + \frac{\beta}{N} \right) - \frac{\delta}{N} - \phi N \right\} \tag{A5}$$

of identical countries. The objective function W has strictly decreasing differences in (N, γ) because for any $\gamma_H > \gamma_L$ and and $N_H > N_L$,

$$W(N_H, \gamma_H) - W(N_H, \gamma_L) = \beta \frac{\gamma_H - \gamma_L}{N_H} < W(N_L, \gamma_H) - W(N_L, \gamma_L) = \beta \frac{\gamma_H - \gamma_L}{N_L}. \tag{A6}$$

Thus, the welfare-maximizing number of countries N_1^* is decreasing in γ in the sense of monotone comparative statics. It is likewise decreasing in β, and increasing in δ and ϕ.

By the same reasoning, the utilitarian welfare optimum with unions is a world union composed of a number

$$N_2^* = \arg\max_{N \in \mathbb{N}} \left\{ -\frac{\delta}{N} - \phi N \right\} \tag{A7}$$

of identical countries.

A.5 Equilibrium Conditions for the General Model

The general model used in this paper is discussed in Section 4.3 when we assume that $\pi \geq 0$ and $\rho \geq 0$. The model of war and conquest of Section 4 applies in the limit as $\rho \to 0$, while the model of diplomacy in Section 3 applies in the limit as $\pi \to 0$ and $\rho \to 0$. In this Appendix, we discuss the different possible equilibria of the general model.

A.5.1 Law and Diplomacy

Assume first that the core decides to forego warfare and join the free world, so $F = 1$. Then we can define the welfare of core localities as:

$$W^C(S_C, U_C) = -\eta + (\gamma + \rho\pi)(1 - \beta) + (\gamma U_C + \rho \min\{U_C, \pi\})\beta - \delta S_C - \frac{\phi}{S_C} - \kappa I_C^U \tag{A8}$$

and the welfare of periphery localities as:

$$W^P(S_P, U_P) = -\eta + \gamma(1 - \beta + \beta U_P) - \delta S_P - \frac{\phi}{S_P} - \kappa I_P^U, \tag{A9}$$

where I_C^U and I_P^U are indicator functions that take value 1 if $S_C = U_C$ and $S_P = U_P$, respectively, and zero otherwise. Utilitarian social welfare for the entire world equals:

$$W\left(S_C, S_P, U_C, U_P\right) = \pi W^C\left(S_C, U_C\right) + \left(1-\pi\right) W^P\left(S_P, U_P\right). \tag{A10}$$

There are three possible equilibrium political structures. The first is a single-level governance structure with countries of optimal sizes

$$S_{1C}^* = \sqrt{\frac{\phi}{\delta - \beta\left(\gamma+\rho\right)}} \text{ and } S_{1P}^* = \sqrt{\frac{\phi}{\delta - \beta\gamma}}, \tag{A11}$$

and without any unions.[33] Utilitarian world welfare under this first structure is:

$$\begin{aligned} W^1 &\equiv W\left(S_{1C}^*, S_{1P}^*, S_{1C}^*, S_{1P}^*\right) \\ &= -\eta + \left(\gamma + \rho\pi^2\right)\left(1-\beta\right) - 2\left\{\pi\sqrt{\phi\left[\delta - \beta\left(\gamma+\rho\right)\right]} + \left(1-\pi\right)\sqrt{\phi\left(\delta-\beta\gamma\right)}\right\}. \end{aligned} \tag{A12}$$

The second potential equilibrium is a two-level governance structure with countries of optimal size

$$S_{2C}^* = S_{2P}^* = \sqrt{\frac{\phi}{\delta}} \tag{A13}$$

and with a world union: $U_C = U_P = 1$. Utilitarian world welfare under this second structure is:

$$W^2 \equiv W\left(S_{2C}^*, S_{2P}^*, 1, 1\right) = -\eta + \left(\gamma + \rho\pi^2\right) - 2\sqrt{\phi\delta} - \kappa. \tag{A14}$$

The third potential equilibrium features a two-level governance structure for the core with countries of size $S_{3C}^* = S_{2C}^*$ and a core union $U_C = \pi$, but a single-level governance structure for the periphery with countries of size $S_{3P}^* = S_{1P}^*$ and no union ($U_P = S_{3P}^*$). Utilitarian world welfare under this third structure is:

$$\begin{aligned} W^2 &\equiv W\left(S_{2C}^*, S_{2P}^*, 1, 1\right) \\ &= -\eta + \left(\gamma + \rho\pi^2\right)\left(1-\beta\right) + \left(\gamma+\rho\right)\pi^2\beta - 2\left[\pi\sqrt{\phi\delta} + \left(1-\pi\right)\sqrt{\phi\left(\delta-\beta\gamma\right)}\right] - \pi\kappa. \end{aligned} \tag{A15}$$

The equilibrium political structure under diplomacy is the one that delivers the highest

[33] Equation (A11) assumes that $\delta > \phi\pi^{-2} + \beta\left(\eta+\rho\right)$, so that there is enough preference heterogeneity to ensure that the optimal core country size is always smaller than the entire core.

welfare:

$$\arg\max W \ni \begin{cases} (S_{1C}^*, S_{1P}^*, S_{1C}^*, S_{1P}^*) & \text{if } W^1 \geq \max\{W^2, W^3\} \\ (S_{2C}^*, S_{2P}^*, 1, 1) & \text{if } W^2 \geq \max\{W^1, W^3\} \\ (S_{3C}^*, S_{3P}^*, \pi, S_{3P}^*) & \text{if } W^3 \geq \max\{W^1, W^2\}. \end{cases} \tag{A16}$$

The core union yields higher welfare than single-level governance ($W^3 > W^1$) if:

$$\kappa < \bar{\kappa}_1(\gamma) \equiv \beta(\gamma + \rho)\pi - 2\sqrt{\phi}\left[\sqrt{\delta} - \sqrt{\delta - \beta(\gamma + \rho)}\right], \tag{A17}$$

for an increasing and concave threshold ($\partial\bar{\kappa}_1/\partial\gamma > 0 > \partial^2\bar{\kappa}_1/\partial\gamma^2$) such that $\partial\bar{\kappa}_1/\partial\beta > 0$, $\partial\bar{\kappa}_1/\partial\delta > 0$, $\partial\bar{\kappa}_1/\partial\phi < 0$, $\partial\bar{\kappa}_1/\partial\pi > 0$ and $\partial\bar{\kappa}_1/\partial\rho > 0$.

The core union yields higher welfare than the world union ($W^3 > W^2$) if:

$$\kappa > \bar{\kappa}_2(\gamma) \equiv \beta\gamma(1 + \pi) - 2\sqrt{\phi}\left(\sqrt{\delta} - \sqrt{\delta - \beta\gamma}\right), \tag{A18}$$

for an increasing and concave threshold ($\partial\bar{\kappa}_2/\partial\gamma > 0 > \partial^2\bar{\kappa}_2/\partial\gamma^2$) such that $\partial\bar{\kappa}_2/\partial\beta > 0$, $\partial\bar{\kappa}_2/\partial\delta > 0$, $\partial\bar{\kappa}_2/\partial\phi < 0$, $\partial\bar{\kappa}_1/\partial\pi > 0$ and $\partial\bar{\kappa}_2/\partial\rho = 0$.

Single-level governance yields higher welfare than the world union ($W^1 > W^2$) if:

$$\kappa > \pi\bar{\kappa}_1 + (1 - \pi)\bar{\kappa}_2. \tag{A19}$$

The two functions $\bar{\kappa}_1(\gamma)$ and $\bar{\kappa}_2(\gamma)$ have a single crossing because

$$\bar{\kappa}_1(0) > \bar{\kappa}_2(0) = 0 \text{ and } \frac{\partial\bar{\kappa}_2}{\partial\gamma} > \frac{\partial\bar{\kappa}_1}{\partial\gamma}. \tag{A20}$$

In other words, the core union can follow but not precede single-level governance and precede but not follow the world union because

$$\frac{\partial}{\partial\gamma}\left(W^2 - W^3\right) = \beta\left[1 - \pi^2 - (1 - \pi)S_{3P}^*\right] > 0 \tag{A21}$$

and

$$\frac{\partial}{\partial\gamma}\left(W^3 - W^1\right) = \beta\pi(\pi - S_{1C}^*) > 0. \tag{A22}$$

A.5.2 War and Conquest

If there are empires, the analysis is as essentially as it was in Section 4 for $\rho = 0$. The welfare of core localities that form an imperial metropolis is given by:

$$W_l = W^E(E) = -\eta + (\gamma + \rho\pi)(1-\beta) + (\gamma + \rho\mu)\beta E - \delta\mu E - \frac{\phi}{E} - \omega. \tag{A23}$$

The optimal size of empires is larger because so are gains from trade within the metropolis:

$$E^* = \sqrt{\frac{\phi}{\delta\mu - \beta(\gamma + \rho\mu)}}. \tag{A24}$$

Thus, if core localities build empires their welfare is given by:

$$W^E(E^*) = -\eta + (\gamma + \rho\pi)(1-\beta) - 2\sqrt{\phi\left[\delta\mu - \beta(\gamma + \rho\mu)\right]} - \omega. \tag{A25}$$

The free world contains a measure $F < 1 - \pi$ of localities in the periphery. In this case, all the analysis in Section 4 applies and, in particular, Condition (19) still determines whether the free world has a single or two-level governance structure.

When are empires formed? In the absence of empires, the welfare of core localities in the welfare-maximizing political structure is given by:

$$W_C^F = \begin{cases} W^C(S_{1C}^*, S_{1C}^*) & \text{if} \quad W^1 > \max\{W^2, W^3\} \\ W^C(S_2^*, \pi) & \text{if} \quad W^2 > \max\{W^1, W^3\} \\ W^C(S_2^*, 1) & \text{if} \quad W^3 \geq \max\{W^1, W^2\}. \end{cases} \tag{A26}$$

If $W^E(E^*) < W_C^F$, there are no empires, diplomacy prevails and the size of the free world is $F = 1$. If instead $W^E(E^*) \geq W_C^F$, there are $\pi/\mu E^*$ empires of size E^*, and the size of the free world is reduced to $F = 1 - \pi/\mu$.

Core localities prefer empires to peaceful countries ($W^E(E^*) > W^C(S_{1C}^*, S_{1C}^*)$) if:

$$\gamma > \bar{\gamma}_E \text{ such that } 2\sqrt{\phi}\left[\sqrt{\delta - \beta(\bar{\gamma}_E + \rho)} - \sqrt{\delta\mu - \beta(\bar{\gamma}_E + \rho\mu)}\right] = \omega, \tag{A27}$$

with $\partial\bar{\gamma}_E/\partial\beta < 0$, $\partial\bar{\gamma}_E/\partial\phi < 0$, $\partial\bar{\gamma}_E/\partial\mu > 0$, $\partial\bar{\gamma}_E/\partial\omega > 0$ and $\partial\bar{\gamma}_E/\partial\pi = 0$.

Core localities prefer empires to the peaceful core union ($W^E(E^*) > W^C(S_2^*, \pi)$) if:

$$\kappa > \bar{\kappa}_{E\pi}(\gamma) \equiv \omega + \beta(\gamma + \rho)\pi - 2\sqrt{\phi}\left[\sqrt{\delta} - \sqrt{\delta\mu - \beta(\gamma + \rho\mu)}\right], \tag{A28}$$

for a concave threshold ($\partial^2 \bar{\kappa}_{E\pi}/\partial\gamma^2 < 0$) with

$$\bar{\kappa}_{E\pi}\left(\bar{\gamma}_E\right) = \bar{\kappa}_1\left(\bar{\gamma}_E\right) \text{ and } \frac{\partial \bar{\kappa}_{E\pi}}{\partial \gamma} < \frac{\partial \bar{\kappa}_1}{\partial \gamma}, \tag{A29}$$

and such that $\partial \bar{\kappa}_{E\pi}/\partial\phi < 0$, $\partial \bar{\kappa}_{E\pi}/\partial\mu > 0$, $\partial \bar{\kappa}_{E\pi}/\partial\omega = 1$, $\partial \bar{\kappa}_{E\pi}/\partial\pi > 0$ and $\partial \bar{\kappa}_{E\pi}/\partial\rho > 0$.

Core localities prefer empires to the peaceful world union ($W^E\left(E^*\right) > W^C\left(S_2^*, 1\right)$) if:

$$\kappa > \bar{\kappa}_{E1}\left(\gamma\right) \equiv \omega + \beta\left(\gamma + \rho\pi\right) - 2\sqrt{\phi}\left[\sqrt{\delta} - \sqrt{\delta\mu - \beta\left(\gamma + \rho\mu\right)}\right], \tag{A30}$$

for an increasing and concave threshold ($\partial \bar{\kappa}_{E1}/\partial\gamma > 0 > \partial^2 \bar{\kappa}_{E1}/\partial\gamma^2$) with

$$\bar{\kappa}_{E1}\left(\gamma\right) > \bar{\kappa}_{E\pi}\left(\gamma\right) \text{ and } \frac{\partial \bar{\kappa}_{E\pi}}{\partial \gamma} < \frac{\partial \bar{\kappa}_{E1}}{\partial \gamma} < \pi\frac{\partial \bar{\kappa}_1}{\partial \gamma} + \left(1 - \pi\right)\frac{\partial \bar{\kappa}_2}{\partial \gamma}, \tag{A31}$$

and such that $\partial \bar{\kappa}_{E1}/\partial\phi < 0$, $\partial \bar{\kappa}_{E1}/\partial\mu > 0$, $\partial \bar{\kappa}_{E1}/\partial\omega = 1$, $\partial \bar{\kappa}_{E1}/\partial\pi > 0$ and $\partial \bar{\kappa}_{E1}/\partial\rho > 0$.

References

[1] Anderson, James E. and Eric van Wincoop (2004). "Trade Costs," *Journal of Economic Literature* 42(3): 691–751.

[2] Andrews, Rhys, and George Boyne (2009). "Size, Structure and Administrative Overheads: An Empirical Analysis of English Local Authorities," *Urban Studies* 46(4): 739–759.

[3] Berry, Christopher B. (2009). *Imperfect Union: Representation and Taxation in Multilevel Governments*. Cambridge: Cambridge University Press.

[4] Blom-Hansen, Jens, Kurt Houlberg and Søren Serritzlew (2014). "Size, Democracy, and the Economic Costs of Running the Political System," *American Journal of Political Science* 58(4): 790–803.

[5] Blom-Hansen, Jens, Kurt Houlberg, Søren Serritzlew and Daniel Treisman (2016). "Jurisdiction Size and Local Government Policy Expenditure: Assessing the Effect of Municipal Amalgamation," *American Political Science Review* 110(4): 812–831.

[6] Blesse, Sebastian, and Thushyanthan Baskaran (2016). "Do Municipal Mergers Reduce Costs? Evidence from a German Federal State," *Regional Science and Urban Economics* 59: 54–74.

[7] Dahl, Robert A., and Edward R. Tufte (1973). *Size and Democracy*. Stanford, CA: Stanford University Press.

[8] Donaldson, Dave (2018). "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review* 108(4–5): 899–934.

[9] Dutt, Pushan, Ilian Mihov and Timothy Van Zandt (2013). "The Effect of WTO on the Extensive and the Intensive Margins of Trade." *Journal of International Economics* 91(2): 204–219.

[10] Fan, C. Simon, Chen Lin, and Daniel Treisman (2009). "Political Decentralization and Corruption: Evidence from Around the World," *Journal of Public Economics* 93(1–2): 14–34.

[11] Head, Keith and Thierry Mayer (2014). "Gravity Equations: Workhorse,Toolkit, and Cookbook," in *Handbook of International Economics*, Vol. 4, edited by Gita Gopinath, Elhanan Helpman and Kenneth Rogoff, 131–195. Amsterdam: North Holland.

[12] Helpman, Elhanan, Marc Melitz and Yona Rubinstein (2008). "Estimating Trade Flows: Trading Partners and Trading Volumes," *Quarterly Journal of Economics*, 123(2): 441–487.

[13] Hummels, David (2007). "Transportation Costs and International Trade in the Second Era of Globalization", *Journal of Economic Perspectives* 21(3): 131–154.

[14] Lassen, David D., and Søren Serritzlew (2011). "Jurisdiction Size and Local Democracy: Evidence on Internal Political Efficacy from Large-Scale Municipal Reform," *American Political Science Review* 105(2): 238–258.

[15] Le Galès, Patrick, and Peter John (1997). "Is the Grass Greener on the Other Side? What Went Wrong with French Regions, and the Implications for England," *Policy & Politics* 25(1): 51–60.

[16] Levinson, Marc (2006). *The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger*. Princeton, NJ: Princeton University Press.

[17] McCallum, John (1995). "National Borders Matter: Canada-U.S. Regional Trade Patterns," *American Economic Review* 85(3): 615–623

[18] Musgrave, Richard A. (1971). "Economics of Fiscal Federalism," *Nebraska Journal of Economics and Business* 10(4): 3–13.

[19] Pascali, Luigi (2017) "The Wind of Change: Maritime Technology, Trade and Economic Development," *American Economic Review* 107(9): 2821–2854.

[20] Reingewertz, Yaniv (2012). "Do Municipal Amalgamations Work? Evidence from Municipalities in Israel," *Journal of Urban Economics* 72(2–3): 240–251.

[21] Steinwender, Claudia (2018). "Real Effects of Information Frictions: When the States and the Kingdom became United," *American Economic Review* 108(3): 657–696.

[22] Strumpf, Koleman S., and Felix Oberholzer-Gee (2002). "Endogenous Policy Decentralization: Testing the Central Tenet of Economic Federalism," *Journal of Political Economy* 110(1): 1–36.