Unified China and Divided Europe

Chiu Yu Ko, Mark Koyama, Tuan-Hwee Sng*

May 2015

Abstract

This paper studies the causes and consequences of political centralization and fragmentation in China and Europe. We build a model to argue that a severe and unidirectional threat of external invasion fostered political centralization in China while Europe faced a wider variety of smaller external threats and remained politically fragmented. We test our hypothesis using data on the frequency of nomadic attacks and political unification in China. Our model allows us to explore the economic consequences of political centralization and fragmentation. Political centralization in China led to lower taxation and hence faster population growth during peacetime than in Europe. But it also meant that China was relatively fragile in the event of an external invasion. Our results are consistent with historical evidence of warfare, capital city location, tax levels, and population growth in both China and Europe.

Keywords: China; Europe; Great Divergence; Political Fragmentation; Political Centralization

JEL Codes: H2, H4, H56; N30; N33; N35; N40; N43; N45

*Chiu Yu Ko and Tuan-Hwee Sng, Department of Economics, National University of Singapore. Mark Koyama, Department of Economics, George Mason University. Email: kochiuyu@nus.edu.sg, mkoyama2@gmu.edu, tsng@nus.edu.sg.
This paper has benefited from seminar presentations at Academia Sinica, Shandong University, Shanghai University of Finance and Economics, Chinese University of Hong Kong, Hitotsubashi University, American University, Hong Kong University, the Osaka Workshop on Economics of Institutions and Organizations, the Washington Area Economic History Workshop at George Mason University, the 2014 EHS Annual Conference, ISNIE 2014, the 2014 EHA Annual Meeting, and the 2015 ASSA meetings. We are grateful for comments from Warren Anderson, Christophe Chamley, Jiahua Che, Mark Dincecco, Rue Esteves, James Fenske, Nicola Gennaioli, Noel Johnson, Debin Ma, Andrea Matranga, Joel Mokyr, Nancy Qian, Jared Rubin, Noah Smith, Yannay Spitzer, Alex Tabarrok, Zhigang Tao, Denis Tkachenko, Melanie Meng Xue, Se Yan, Helen Yang, and many others. We are grateful to Peter Brecke for sharing his Conflict Catalog Dataset and to Ryan Budny, Pei Zhi Chia, and Jane Perry for research support.

1 Introduction

Since Montesquieu, scholars have attributed Europe's success to its political fragmentation (Montesquieu, 1989; Jones, 2003; Mokyr, 1990; Diamond, 1997). Nevertheless, throughout much of history, the most economically developed region of the world was China, which was typically a unified empire. This contrast poses a puzzle that has important implications for our understanding of the origins of modern economic growth: Why was Europe perennially fragmented after the collapse of Rome? Why was political centralization an equilibrium for most of Chinese history? Can this fundamental difference in political institutions account for important disparities in Chinese and European growth patterns?

This paper proposes a unified framework to (a) provide an explanation for the different political equilibria in China and Europe; (b) explore the economic consequences of political centralization and fragmentation.

Our model predicts when and where empires are viable based on the nature and intensity of the external threats that they face. We focus on the interaction of geography and external threats in determining China's recurring unification and Europe's enduring fragmentation. Historically, China faced a severe, unidirectional threat from the Eurasian steppe or grassland. Europe confronted several smaller threats from Scandinavia, Central Asia, the Middle East, and North Africa. The inefficiency of multitasking means that empires tended not be viable in Europe and political fragmentation was the norm. On the other hand, empires were more likely to emerge and survive in China because the nomadic threat threatened the survival of small states more than larger ones.

The different equilibria that we identify had important economic consequences. Political centralization allowed China to avoid wasteful interstate competition and to tax lightly. This enabled it to enjoy more rapid economic and population growth during peacetime. Taxes were higher in Europe than in China. However, the presence of multiple states to protect different parts of the continent meant that Europe was relatively more robust to both known threats and unexpected negative shocks.

To test the mechanisms identified in our model, we use time series analysis to show that an increase in the frequency of nomadic attacks on China is associated with more political centralization in historical China. Our estimates suggest that each additional nomadic attack per decade increased the probability of political unification in China by about 6.3% in the long run. We also use our theory in conjunction with narrative and qualitative evidence to discuss why no stable European empire emerged after the collapse of Rome. Furthermore, our model offers an explanation as to why Chinese rulers tended to locate their capitals on the periphery of their empire. Finally, we provide evidence supporting the predictions of the model concerning taxation and population growth.

Our paper relates to several strands of literature. Our theoretical framework builds on the research on the size of nations originated by Friedman (1977) and Alesina and Spolaore (1997, 2003). In particular, our emphasis on the importance of external threats is related to the insights of Alesina and Spolaore (2005) who study the role of war in shaping political boundaries. In examining the causes of political fragmentation and centralization in China and Europe, we build on earlier work that points to the role of geography, such as Diamond (1997), and on many historians who stress how the threat of nomadic invasion from the steppe shaped Chinese history (Lattimore, 1940; Grousset, 1970; Huang, 1988; Barfield, 1989; Gat, 2006).

Our model also sheds light on the optimal location of capital cities. Economic theory generally predicts that capital cities should be centrally located to maximize tax revenue or improve governance (e.g., Alesina and Spolaore 2003; Campante et al. 2014). This is confirmed by empirical studies showing that isolated national or subnational capital cities are associated with greater corruption (Olsson and Hansson, 2011; Campante and Do, 2014). We show that if the effectiveness of public goods provision (military defense in our example) differs with the location of provision, it may be optimal to establish the political center of the empire away from its economic or population center. This helps explain why the Romans moved their capital city from Rome to Constantinople in AD 330, and why Beijing, a city on the northern periphery of China proper, was China's political center for more than six centuries.

By developing a new explanation of why Europe was persistently fragmented, we complement the literature that emphasizes the positive economic consequences of European political fragmentation, which include promoting economic and political freedom (Montesquieu, 1989; Pirenne, 1925; Hicks, 1969; Jones, 2003; Hall, 1985; Rosenberg and Birdzell, 1986); encouraging experiments in political structures and investments in state capacity (Baechler, 1975; Cowen, 1990; Tilly, 1990; Hoffman, 2012; Gennaioli and Voth, 2013);[1] intensifying interstate conflicts and thereby promoting urbanization (Voigtländer and Voth, 2013b);[2] and fostering innovation and scientific development (Diamond, 1997; Mokyr, 2007; Lagerlof, 2014).[3] Our analysis is also related to the rise of state capacity in Europe

[1]Baechler observed that 'political anarchy' in Europe gave rise to experimentation in different state forms (Baechler, 1975, 74). Cowen (1990) argues that interstate competition in Europe provided an incentive for early modern states to develop capital markets and pro-market policies. Tilly (1990) studies the role capital-intensive city states played in shaping the emergence of nation states in Europe. Hoffman (2012) uses a tournament model to explain how interstate competition led to military innovation in the early modern Europe. Gennaioli and Voth (2013) show that the military revolution induced investments in state capacity in some, but not all, European states.

[2]Voigtländer and Voth (2013b) argue that political fragmentation interacted with the Black Death so as to shift Europe into a higher income steady-state Malthusian equilibrium.

[3]Diamond argues that 'Europe's geographic balkanization resulted in dozens or hundreds of independent, competing statelets and centers of innovation' whereas in China 'a decision by one despot could and repeatedly did halt innovation' (Diamond, 1997, 414–415). Mokyr notes that 'many of the most influential and innovative intellectuals took advantage of ... the competitive 'states system'. In different ways, Paracelsus, Comenius, Descartes, Hobbes, and Bayle, to name but a few, survived through strategic moves across national boundaries. They were able to flee persecutors, and while this imposed no-doubt considerable hardship, they survived and prospered' (Mokyr, 2007, 24). Lagerlof (2014) develops a growth model that emphasizes the benefits to scale in innovation under political unification and a greater incentive to innovate under political fragmentation. He calibrates the model to the initial conditions of China and Europe and shows that there are parameter values in which political fragmentation can give rise to the emergence of sustained

(Dincecco, 2009; Dincecco and Katz, 2014; Johnson and Koyama, 2013, 2014a,b) and to recent research that emphasizes other aspects of Europe's possible advantages in the Great Divergence such as the higher age of first marriage than the rest of the world (Voigtländer and Voth, 2013a); public provision of poor relief versus reliance on clans as was the case in China (Greif et al., 2012); institutions that were less reliant on religion (Rubin, 2011); greater human capital (Kelly et al., 2013); and higher social status for entrepreneurs and inventors (McCloskey, 2010).

Perhaps the closest argument to ours is that of Rosenthal and Wong (2011) who argue that political fragmentation led to more frequent warfare in medieval and early modern Europe, which imposed high costs but also lent an urban bias to the development of manufacturing which led to more capital intensive forms of production. Like them, we emphasize that political fragmentation was costly for Europe, but we develop a different argument based on the observation that the costs of political collapse and external invasion were particularly high in China. Theoretically and empirically, we show that the Chinese empire could indeed have been more conducive to Smithian economic expansion during stable periods as Rosenthal and Wong claim, but we also note that it was less robust to external shocks, and this greater volatility of population and economic output was a major barrier to sustained economic growth in China before 1800.

The rest of the paper is structured as follows. Section 2 provides historical evidence that characterizes (i) the extent to which China was politically unified and Europe fragmented throughout their respective histories, and (ii) the degree to which both China and Europe were threatened by external invasions. In Section 3 we introduce a model of political centralization and decentralization. Section 4 provides empirical evidence to support our hypothesis that a severe threat from the Eurasian steppe discouraged political fragmentation in China. In Section 5, we show that our model provides a coherent framework that can help to make sense of the choice of capital cities, differential levels of taxation, and population growth patterns in historical China and Europe. Section 6 concludes.

2 The Puzzle: Unified China and Divided Europe

Unit of analysis. States and state systems first emerged in areas suitable for settled agriculture where food surpluses were available to form the basis of taxation (Childe, 1936; Carneiro, 1970). In this paper, we focus on the two continuous agricultural zones at both ends of Eurasia, China and Europe (Figure 1). For Europe, we focus on its western portion, or the area west of the Hajnal line (Hajnal, 1965).[4] Meanwhile, we equate China with China proper, an area bounded by the

growth in Europe.

[4]Our analysis is unchanged if we consider instead the Ural mountains as the eastern boundary of 'Europe.' Indeed our framework provides a potential explanation as to why empires were more frequent in Eastern Europe than in Western Europe.

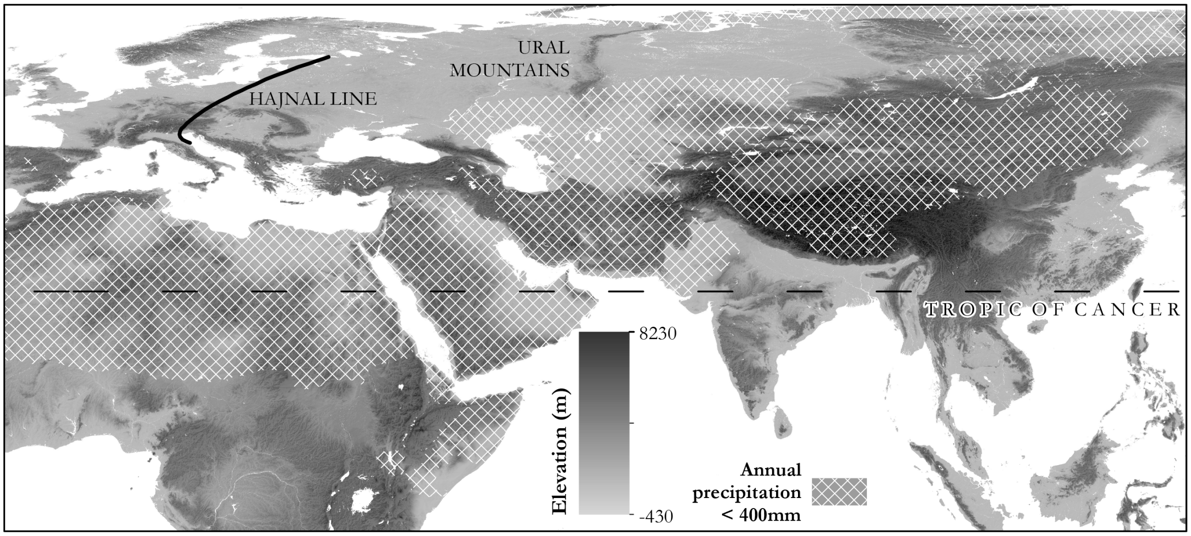


Figure 1: Two ends of Eurasia: Western Europe (i.e., west of the Hajnal line) and China proper (i.e., the agricultural zone bounded by the 400mm isohyet line in the north, the Himalayas and other mountain ranges in the west, tropical rainforests in the south, and the Pacific Ocean in the east). Notice that China was relatively isolated except for its northern frontier. By contrast, Europe was connected to the rest of Eurasia and Africa in multiple directions.

Pacific Ocean to its east, the thick tropical rainforests of Indochina to its south, huge mountain ranges—including the Himalayas—to its west, and the Great Wall to its north. Although the Great Wall was manmade, it overlaps largely with the 400mm isohyet line, which approximates the northern limit of rainfall agriculture (Food and Agriculture Organization, 1986). In other words, the Great Wall delineates the ecological divide between the steppe nomads of Central Asia and the agricultural population in the river basins of China.[5] ‘China’ and ‘Europe’ are comparable in size: China proper covers a land area of 2.8 million square kilometers, while Western Europe has slightly more than 2.5 million square kilometers.

Patterns of unification and fragmentation. Chinese historical records indicate that less than 80 states ruled over parts or all of China between AD 0 and 1800 (Wilkinson, 2012). Nussli (2011) provides data on the sovereign states in existence at hundred year intervals in Europe. Figure 2 plots the number of sovereign states in China and in Europe for the preindustrial period. There have always been more states in Europe than in China throughout the past two millennia; in fact since the Middle Ages there have been an order of magnitude more states in Europe than in China.[6]

[5]As Lattimore (1940, 25) put it, “[The Great Wall] is the product of social emphasis continuously applied along a line of cleavage between environments.”

[6]The Nussli (2011) data does not capture all political entities in Europe since that number is unknown—there may have been as many as 1000 sovereign states within the Holy Roman Empire alone—but it does record the majority of large and small political entities (Abramson, 2013). By contrast, the Chinese dynastic tables are well known and the potential for disagreement is immaterial for our purposes. We count only sovereign states. Including vassal states

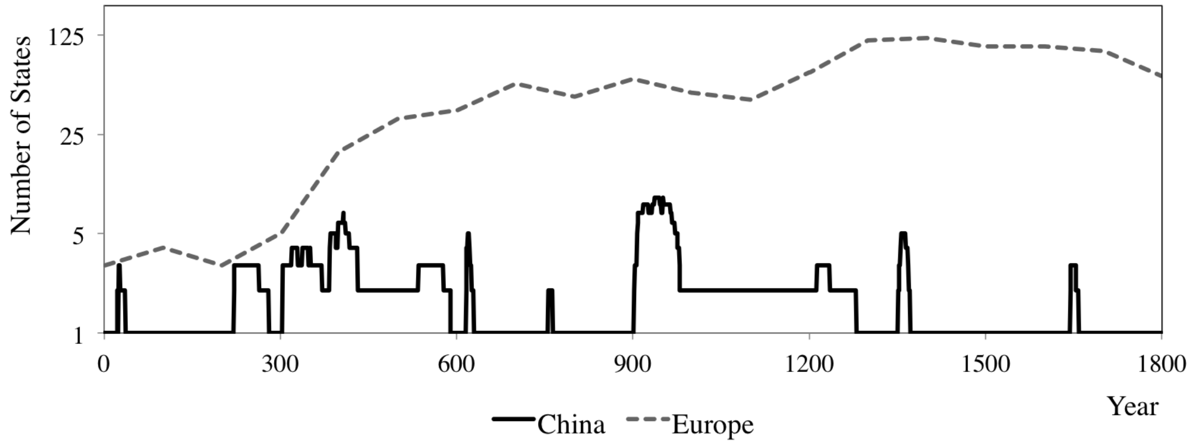


Figure 2: Number of sovereign states in China and Europe, 1–1800 (Nussli, 2011; Wei, 2011).

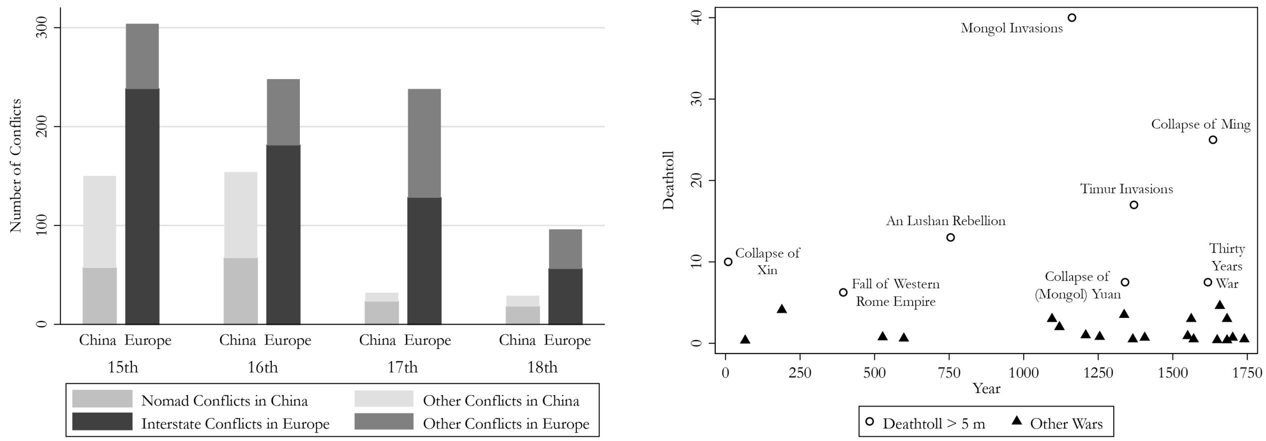
The Chinese first established a unitary empire in the third century BC, before Rome's dominance of the Mediterranean (Elvin, 1973; Needham, 1995; Fukuyama, 2011). Moreover, the Chinese empire outlasted Rome. Although individual dynasties rose and fell, China as an empire survived until 1912. Between AD 1 and 1800, the landmass between the Mongolian steppe and the South China Sea was ruled by one single authority for 1007 years (Ko and Sng, 2013). Every period of political disunity was followed by reunification—in China, Humpty Dumpty could always be put back together again. This phenomenon is captured in the famous opening lines of the classic Chinese novel *Romance of the Three Kingdoms*: 'The empire, long divided, must unite; long united, must divide. Thus it has ever been.'

In comparison, Europe after the fall of the Roman Empire was characterized by persistent political fragmentation—no subsequent empire was able to unify a large part of the continent for more than a few decades. The number of states in Europe increased from 37 in 600 AD to 61 in 900 and by 1300 there were 114 independent political entities. The level of political fragmentation in Europe remained high during the early modern period.

Patterns of Warfare. It is therefore unsurprising that interstate warfare, or military conflicts between sedentary societies, were much more common in Europe than in China. Figure 3a, which is derived from Peter Brecke's Conflict Catalog Dataset (Brecke, 1999), shows that intense interstate competition contributed to Europe's higher overall frequency of warfare between 1400 and 1800.

By contrast, a majority of China's military conflicts were with nomads from the Eurasian steppe. Steppe nomads regularly invaded Chinese regions in the past two millennia. According to Chaliand (2005), out of the seven major waves of nomadic invasions witnessed in Eurasia since the first century, China was involved in six occasions, while Europe was affected only twice (See Appendix A.1).

would further strengthen the argument.



(a) Number of Violent Conflicts (Brecke, 1999) (b) Largest Wars By No. of Deaths (White, 2013)

Figure 3: The Nature, Frequency, and Intensity of Warfare in China and Europe

Figure 3b provides another intriguing—and hitherto overlooked—observation: the most violent wars of the preindustrial period occurred in Asia and particularly in China. While warfare was less common in China, it was more costly than in Europe.[7] Only two wars with estimated death tolls in excess of 5 million are recorded for Europe compared with five for China.[8] Wars in China such as the An Lushan Rebellion, the Mongol invasions, and the Ming-Manchu transition were extremely costly because they involved the collapse or near collapse of entire empires. Notably, each of these wars had a nomadic dimension.[9] By contrast, warfare in Europe was endemic, but rarely resulted in large scale socio-economic collapse. The only European war that matched the death tolls of the worst conflicts in Chinese history was the Thirty Years War, which was a German civil war into which the other European powers were drawn.

We argue that the patterns in Figures 2 and 3 are connected: while the immediate effect of a nomadic invasion was to create chaos and weaken sedentary regimes, in the long run the presence of a prominent and irresolute steppe threat along China's northern border constituted a centripetal force that regularly pressed the constituent regions of China toward unification; meanwhile, the foremost concerns of European regions were the idiosyncratic threats and problems that they individually faced,

[7] All data on deaths from warfare in the preindustrial period are highly speculative, but for our purposes what is important is the order of magnitude rather than the precise numbers reported. Unfortunately, it is not possible to calculate per capita death tolls for many of these conflicts due to the absence of reliable population data for the relevant area affected by the conflict.

[8] The majority of deaths in preindustrial wars did not occur in the battlefield but were the result of disease and pressure on food supplies (see Voigtländer and Voth 2013b, 781 for a discussion).

[9] The Mongols and Manchu were nomadic or semi-nomadic. An Lushan was a general of nomadic origins (See Section 5.1). The collapse of the Xin dynasty (9–23 AD) was a civil war that began with a costly military campaign against the nomads that—coupled with massive flooding along the Yellow River—led to widespread rebellion in China (China's Military History Editorial Committee, 2003).

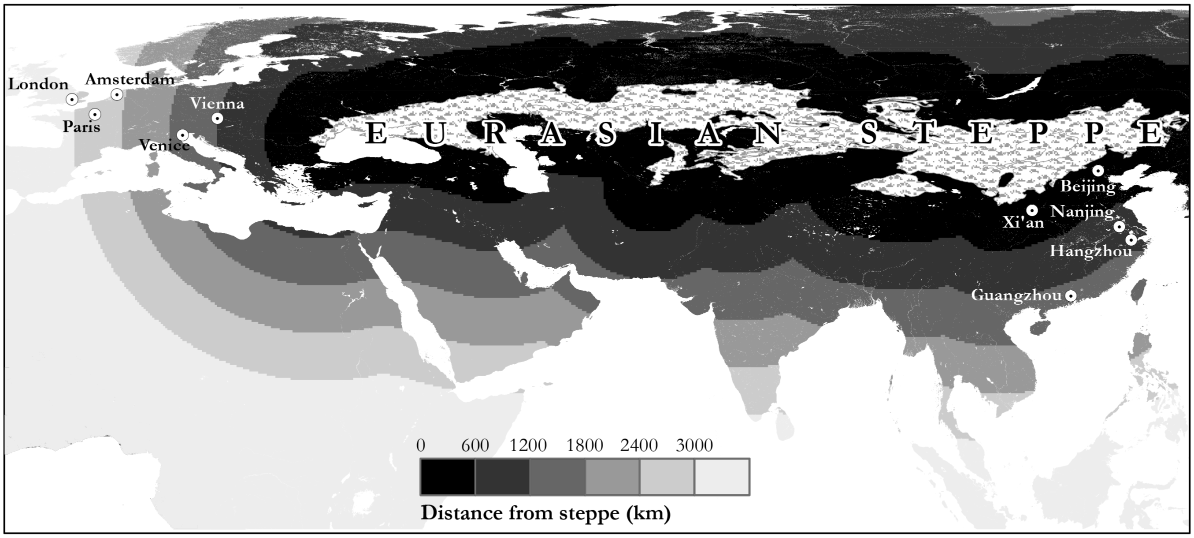


Figure 4: The Eurasian Steppe and Major Cities in China and Europe. Each shade represents 600 kilometers from the steppe.

which in turn discouraged the rise of empires in Europe. To be sure, the presence of an overarching security threat (of lack thereof) was not the only factor that mattered—there were other forces, both centripetal and centrifugal, at work—but it increased the likelihood of Chinese unification and European fragmentation significantly.

The Eurasian Steppe Throughout its history, China was repeatedly invaded by the nomadic and semi-nomadic people north of its borders: Hu, Xiongnu, Xianbei, Juanjuan, Uygurs, Khitan, Jurchen, Mongols, and Manchus (Grousset, 1970; Barfield, 1989; Di Cosmo, 2002a; Chaliand, 2005). This was an inevitable outcome of China's proximity to the grasslands of Central Asia. Figure 4 illustrates the distance of cities in China and Europe from the Eurasian steppe. As it makes clear, Guangzhou, the southernmost major Chinese city, was almost as close to the steppe as Vienna, the easternmost major western European city.

According to Lattimore (1940), the struggle between the pastoral herders in the steppe and the settled populations in China was first and foremost an ecological one.[10] The geography of Eurasia created a natural divide between the river basins of China and the Eurasian steppe. In the Chinese river basins, fertile alluvial soil, sufficient rainfall, and moderate temperature encouraged the early development of intensive agriculture. In the steppe, pastoralism emerged as an adaptation to the

[10] Lieberman (2009) distinguishes between China, which lies in the exposed zones of inner Asia, and the protected rimland of Europe and Southeast Asia. He notes that 'For centuries nonpareil equestrian skills, an ethos focused on hunting and warfare, proficiency with the short double-reflex bow (which allowed volleys from horseback), tactical flexibility, a ruthlessness and stamina demanded by an unforgiving environment, remarkable mobility, and a far larger percentage of men trained for war than in settled societies' meant that settled society faced a perennial threat from the horsemen of the steppe.

arid environment. Given the fragile ecology of the steppe, where droughts often led to extensive and catastrophic deaths among animal herds, the steppe nomads were impelled to invade their settled neighbors for food during periods of cold temperature (Bai and Kung, 2011; Zhang et al., 2015).

Three interrelated characteristics of the recurring conflicts between the steppe inhabitants and the agrarian Chinese differentiate them from the typical interstate wars. First, as observed by Central Asian specialists (Lattimore, 1940; Barfield, 1989) and demonstrated empirically by Bai and Kung (2011) and Zhang et al. (2015), nomadic-agrarian conflicts were often climate driven and therefore largely exogenous.

Second, warfare between the steppe inhabitants and the agrarian Chinese was asymmetric in ways that favored the nomads. Although the sedentary Chinese were more populous by far, before the advent of effective gunpowder weapons, the expertise of the steppe nomads on horseback allowed them to develop mobile and powerful cavalry units that could easily outflank and outmaneuver infantry-based armies (Barfield, 1989; Graff, 2002; Gat, 2006).

Importantly, horses were a location-specific asset. Horses bred in the steppe were hardy and had great vigor as they were raised in an environment similar to that of wild horses (Food and Agriculture Organization, 1984).[11] Meanwhile, agrarian China lacked good quality horses and skilled horsemen due to inadequate pastures. Indeed, steppe confederacies that invaded and settled inside China usually lost their distinct military advantages once they became sedentary and sinicized. Interestingly enough, they often found it necessary to rebuild the Great Wall to defend themselves against new nomadic powers that filled the vacuum that they left behind on the steppe.[12]

The third characteristic that sets nomadic-agrarian conflicts apart from typical interstate wars is the absence of towns and cities in the steppe for the sedentary people to capture in times of war. Since the main properties of the steppe pastoralists were their animal herds, which could be moved readily, nomads needed not defend their land against the enemy. When the odds were not in their favor, they could simply retreat into the safe haven of the steppe, where the undifferentiated 'highway of grass' allowed them to reach the Black Sea from Mongolia in a matter of weeks (Frachetti, 2008, 7). In other words, the nomads enjoyed an 'indefinite margin of retreat'—no matter how badly they were defeated in battle, they could never be conquered in war (Lattimore, 1940).

The ineffectiveness of offensives against the nomads was long recognized by the Chinese. In the 3rd century BC, Li Si, the prime minister of the First Emperor, remarked that, 'The Xiongnu have

[11]As Lattimore (1929, 131) observed, 'It takes the sweet grass of rocky hills to bring out the best in Mongol breeds of pony. Those of sandy districts are not only coarse in build, but have wide, spongy hoofs very different from the clean, hard hoofs and springy pasterns of the hill ponies.'

[12]For example, the Northern Wei, descendants of the Xianbei tribes, rebuilt the Great Wall in the 5th century to halt the advance of the Juanjuan. The Northern Zhou did likewise in the 6th century to thwart the Turks. Similar endeavors were made by the Khitans in the 11th century (against the Jurchens) and the Jurchens in the 12th century (against the Mongols).

no fixed cities or forts and no stores of provisions or grain. They move from place to place like flocks of birds and are just as difficult to catch and control. Now if we send parties of lightly equipped soldiers deep into their territory, our men will soon run out of food, and if we try to send provisions after them, the baggage trains will never reach them in time' (Watson, 1971, 194).

Until Russia's expansion into Central Asia in the seventeenth and eighteenth centuries denied the nomads their traditional escape route, the steppe threat was a recurring problem that the Chinese could not permanently resolve (Perdue, 2005). Their best hope for security was the successful containment of the nomadic threat—hence the construction of the first Great Wall immediately after the first unification of China under the Qin dynasty in 221 BC.[13] The project was repeated time and again by successive dynasties at great cost to keep the 'barbarians' at bay.

In other words, the irresolvability of the steppe threat made nomadic invasions a reoccurring theme in Chinese history. Of the ten dynasties that ruled a unified or mostly unified China, three fell to nomadic invaders (Jin, Song, and Ming), of which two were replaced by nomadic dynasties (Yuan and Qing). The fall of three others (Qin, Xin, and Sui) could be traced to wars with their northern neighbors, which placed an unbearable strain on the peasants and led to widespread revolts. The remaining two, Han and Tang, built their golden ages upon the temporary subjugation of the steppe.

Unidirectional versus Multidirectional Threats Many scholars have recognized the importance of the steppe nomads to state formation in ancient China (Lattimore, 1940; Huang, 1988; Lieberman, 2009; Turchin, 2009; Ma, 2012; Deng, 2012). We build on this literature by highlighting another important element in the nature of this threat: the external threats confronting China were unidirectional. There were no major threats from other fronts that would have increased the appeal of a more flexible politically decentralized system.

Before 1800, all major invasions of China came from the north. We argue that this was geographically determined: as Figure 1 illustrates, China was shielded from the south and the west by the Himalayas, the Tibetan plateau, and the dense tropical rain forests of Indochina.[14]

By contrast, Europe's external environment was different in two important ways. First, while Europe was also threatened by invasions from the steppe from Goths, Sarmatians, Vandals, Huns, Avars, Bulgars, Magyars, Pechnegs, Cumans, Mongols, and Turks,[15] the threat was less severe as

[13] During the Warring States period (475–221 BC) when China was divided into about a dozen competing kingdoms, the three that bordered the steppe—Qin, Zhao, and Yan—built long walls that were later linked up to form the first Great Wall of China after Qin successfully unified China.

[14] Lewis (1991, 4–5) notes that the 'high desiccated plateaus and deserts ... dominated the terrain of Mongolia and Turkestan ... formed barriers between it and the Indic and Islamic worlds to the west and south. While it was possible to reach Burma over difficult mountain passes leading there from the upper Yangtze valley, the most practicable routes to the west were by way of Kansu, Inner Mongolia, and eastern Turkestan and then on to Khorasan and southern Russia.'

[15] See Appendix, Table 5, for a list of all major nomadic invasions of both China and Europe.

Western Europe was relatively protected along its eastern flank due to its forests and mountain ranges, and because it was relatively far from the steppe (Figure 4) and was buffered by the semi-pastoral lands of Hungary and Ukraine (Gat, 2006, 383).[16] It is well established that the military superiority of steppe invaders declined dramatically once they entered Europe due to the absence of the vast tracts of grassland required to maintain the high ratio of horses to men that nomadic armies typically relied upon for their effectiveness (Di Cosmo, 2002b).

Second, Europe was less isolated from the rest of Eurasia and consequently prospective European empires typically faced enemies on multiple fronts: Vikings from the north, Arabs, Berbers, and Turks from the south and south-east, Magyars, Mongols, and others from the east (Appendix Table 6). The security challenges that confronted Europe on multiple fronts were particularly substantial in the first millennium. In Section 4.2, we discuss how this contributed to the collapse of the Roman empire and thwarted the attempts of Rome's successors such as the Carolingian empire to reunify Europe.[17]

Alternative theories. Other scholars have offered theories to explain China's tendency toward unification and Europe's persistent fragmentation, ranging from culture and language (e.g., logographic versus phonogramatic writing systems) to topography (e.g., the presence or absence of internal mountain barriers) and institutions (e.g., the Chinese imperial examination system).[18] Our theory and these existing arguments are not mutually exclusive; they reinforce each other.

We do not claim that our theory alone explain the patterns of political consolidation and fragmentation observed. For example, while the presence of multi-sided threats could shed light on why Europe was fragmented in the Middle Ages (see Section 4.2), it cannot explain why the number of European states actually rose between 1000 and 1400 (as depicted in Figure 2). Instead, our argument complements existing institutional and cultural accounts. Tilly (1990) argued that it was the presence of independent city states along the corridor between southern England and northern Italy that prevented the emergence of large empires in Europe at the end of the middle ages, which

[16]The Hunnic invasions of the fifth century, the Magyar invasions of the ninth century, and the Mongol invasions of the thirteenth century pose partial exceptions to this.

[17]Before 1100 Europe faced invasions from north, south, and east. After 1100 the Vikings were Christianized and integrated into the European state system so we no longer classify them as an external threat. Furthermore, the threat of invasion from the north disappeared. Nevertheless Europe continued to face a potential invasion threat from the south and the east, and after 1300 it faced the threat of invasion from the south-east (Appendix Table 6). It is important to note that what matters for our theory is the existence of a potential threat. Thus, Europe continued to face a potential threat of invasion from North Africa due to geographical proximity throughout this period, although the majority of actual invasions from North Africa occurred before 1300. Portugal and Spain faced a threat from North Africa through the sixteenth and seventeenth centuries, but they typically dealt with it through offensive actions (e.g., Charles's V conquest of Tunis 1535—a response to raids by Hayreddin Barbarossa along the southern Italian coast—and Sebastian I's invasion of Morocco which ended in his defeat at the Battle of Alcácer Quibir in 1578).

[18]See Diamond (1997, 322–333) for a detailed discussion.

in turn gave rise to the emergence of modern nation states by the nineteenth century.[19] However, this important and influential theory does not explain the existence of independent city states in Europe in 1500—which he took as given. Our argument complements his hypothesis by helping to explain why it was possible for independent city states to exist in Europe to begin with.

We argue that what geography provided was a regular impetus that nudged China and Europe toward different paths of political development. This, in turn, fostered cultural and institutional developments that further promoted unification in China and fragmentation in Europe. For example, as Di Cosmo (2002a) points out, the recurring conflicts between the steppe inhabitants and the agrarian Chinese helped forge a common cultural identity and a sense of 'Chineseness' among the early competing states in ancient China.

These regular 'geographical nudges' were not sufficient for an empire to emerge, but rather provided necessary conditions for self-reinforcing patterns of political centralization to be established. To see from another example, Twitchett (1979), Fairbank (1992), and other historians have argued that the classical examination system strengthened political centralization in China by undermining the influence of local aristocratic families. Yet as Fairbank (1992) suggests, the perfection of the examination system as a tool for training submissive bureaucrats took time and involved the efforts of successive unified dynasties. Chance alone could not have produced the rounds of trials and errors required for the Chinese rulers to identify and hone institutions that buttressed centralization and dismantle those that promoted centrifugal tendencies.

3 Model

Building on the preceding discussion, we develop a model to explore the consequences of the severe one-sided threat that China faced in contrast with the weaker multi-sided threat faced by states in Europe. To keep our analysis tractable, we consider a continent, which may represent China or Europe, as a Hoteling's linear city of unit length.[20] The continent faces external threats that can be one or two-sided. In particular, the two-sided threat in this one-dimensional model is analogous to a multi-sided threat in the multi-dimensional real world.

The continent contains one or more political regimes. Each regime (a) chooses its capital city (represented by a point along the linear line), (b) taxes its population, and (c) builds a military to resist the external threat and to compete with other states for territory and population. Our central concern is the fiscal viability of the regime(s)—hence the stability of political centralization or fragmentation—given the nature of external threats that the continent confronts.

[19] Similar arguments are also developed by Spruyt (1994) and Finer (1999).

[20] We refer to both Europe and China as 'continents' for convenience.

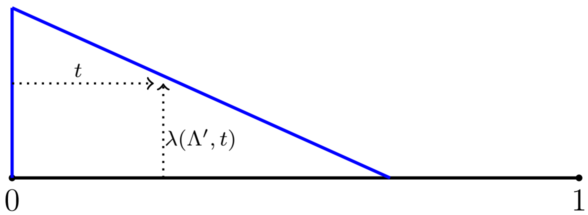


Figure 5: A severe, one-sided threat.

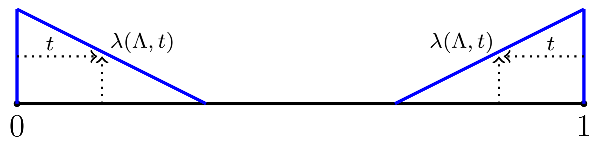


Figure 6: A smaller, two-sided threat.

When setting up the model, we use diagrams (Figures 5–14) to aid illustration. The validity of our results is not tied to the functional forms assumed in the diagrams.

3.1 Setup

We model a continent as a line $[0, 1]$ with a unit mass of individuals uniformly distributed along this line. An individual at $x \in [0, 1]$ is endowed with income $\underline{y} + y$ where y is taxable. For now, we fix the level of taxation at y; we endogenize it later.

The continent faces threats from outside. An external threat of size Λ, if realized, causes gross damage $\lambda(\Lambda, 0) > 0$ at the frontier(s) of the continent. The damage can spread further into the continent: if a point is t distance away from the frontier, the **gross damage** is $\max\{\lambda(\Lambda, t), 0\}$ where $\lambda_1 > 0$, and $\lambda_2 < 0$. The negative derivative of λ with respect to t implies that the threat decreases in strength as it moves inland. We model these external threats as emanating either from one frontier of the continent (at $x = 0$ only without loss of generality) or from both frontiers (Figures 5 and 6). Whether the threat is one-sided or two-sided and the magnitude of Λ depends on the continent's geographical environment, which is exogenously determined.

The continent is divided into $S \in \mathbb{N}_+$ connected, mutually exclusive intervals each ruled by a separate political authority or regime. We take S as given and do not model how regimes arise.[21] Instead, we focus on the fiscal viability of these regimes: we ask, for any given S, is this political configuration stable given the continent's external environment?

In this section, we consider $S \leq 2$. When $S = 1$, there is political centralization and one regime or empire, e, rules the entire continent (i.e., political centralization). When $S = 2$, two regimes, l and r, coexist (i.e., political fragmentation). Regime l is on the left of regime r. For tractability and because we are only interested in analyzing comparable regimes,[22] we treat l and r as identical and focus on the symmetric equilibrium when deriving the results. In the Appendix, we extend the model to study the case when $S > 2$ (A.3). We also provide functional form illustrations for $S = 2$ and $S = 3$ to show that under plausible parametric assumptions, the symmetric equilibrium is in fact the

[21]Historically, the emergence of a regime is often associated with stochastic elements—the birth of a military genius; policy errors made by the incumbent ruler; climatic change; and so on—that are difficult to capture in a model.

[22]If one regime rules a much larger interval than the other one, the continent is effectively politically centralized.

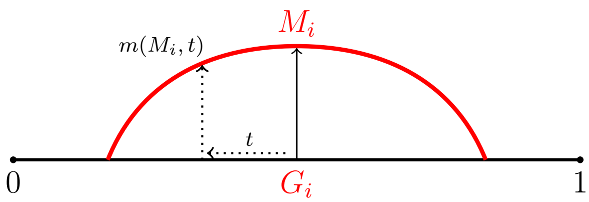


Figure 7: Regime i decides the location of its capital city (G_i) and its military investment (M_i).

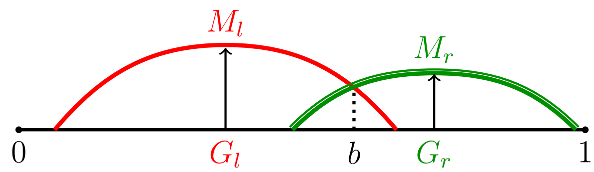


Figure 8: The border (b) between two regimes is determined by the locations of their capital cities (G_l and G_r) and their relative military investments (M_l and M_r).

unique equilibrium (A.4).

A regime may invest in the military to (1) block the external threat, and (2) gain territory (i.e., control a larger interval) when $S = 2$. We assume that the cost of military investment is convex. For regime $i \in \{e, l, r\}$ to provide a military investment of $M_i \geq 0$, it costs $c(M_i)$, which is continuously differentiable and $c(0) = 0$, $c' > 0$, and $c'' > 0$.

A regime's military is strongest at its center of deployment, but its effectiveness deteriorates over distance. Let G_i denote regime i's center of military deployment—referred to here as i's capital city.[23] As illustrated in Figure 7, for a location that is t distance away from G_i, regime i's **military strength** on that location is given by $\max\{m(M_i, t, \beta), 0\}$, where β is a parameter that measures the difficulty of projecting military strength over distance.[24]

Assumption 1 (Military strength)**.** *The function $m\,(M, t, \beta)$ satisfies the following properties:*

1. *(Monotonicity) $m_1 > 0$, $m_2 < 0$, $m_3 \leq 0$ (equality if and only if $t = 0$);*

2. *(Distance effect) $m_{22} < 0$, $m_{12} \leq 0$;*

3. *(Effect of Parameter β) $m_{33} \leq 0$, $m_{13} \leq 0$, $m_{23} \leq 0$.*

Assumption 1 is straightforward. Point 1 states that a regime's military strength is increasing in its military investment, decreasing in distance and also in the difficulty of power projection. Point 2 stipulates that military strength deteriorates over distance at an increasing rate. Furthermore, the marginal effectiveness of military investment is decreasing in distance. Point 3 implies that military

[23]As a matter of interest, the Chinese called the capital *jing-shi*—literally the peak (*jing*) and the military (*shi*).

[24]Historically, the projection of military power was constrained by logistics (van Creveld, 2004). The Roman historian Keith Hopkins highlights the challenges that premodern transportation technologies imposed on political and military control by describing the Roman Empire as 'several months wide—and larger in winter than in summer' (Morris and Scheidel, 2009, 186). According to the Chinese polymath Shen Kuo (1031–1095), during his time a soldier would need one porter for supplies to march 18 days, while three porters would be required for a 31-day campaign (Shen, 2011). Another factor that causes military effectiveness to deteriorate over distance is the cost of controlling armies from afar—in particular, agency problems (See also footnote 42).

strength is decreasing in β at an increasing rate. In addition, a higher β would reduce the marginal effectiveness of military investment and increase the decline in military strength over distance.

We also assume that each regime can only maintain one capital city to reflect the observation that, historically, empires that maintained two or more comparable political-military centers either behaved like multiple states or usually ended up fragmenting into multiple states (see Section 5.1 for a detailed discussion). In Appendix A.5, we show that allowing a regime to set up auxiliary military bases does not affect the results once secession risk is taken into consideration.

As Figure 8 illustrates, when $S = 2$, regime l controls $[0, b]$ and regime r controls $[b, 1]$. The border b depends on the regimes' capital city locations (G_l and G_r) and military investments (M_l and M_r).

Assumption 2 (Border formation)**.** *The border* $b\left(G_l, G_r, M_l, M_r, \beta\right) \in \left[G_l, G_r\right]$ *satisfies:*

1. *(Monotonicity)* $\frac{\partial b}{\partial M_l} > 0$*,* $\frac{\partial b}{\partial M_r} < 0$*,* $\frac{\partial b}{\partial G_l} > 0$*, and* $\frac{\partial b}{\partial G_r} > 0$*;*
2. *(Concavity)* $\frac{\partial^2 b}{\partial M_l^2} \leq 0$*,* $\frac{\partial^2 b}{\partial M_r^2} \geq 0$*,* $\frac{\partial^2 b}{\partial G_l^2} \leq 0$*, and* $\frac{\partial^2 b}{\partial G_r^2} \geq 0$*;*
3. *(Symmetry) When* $G_l = 1 - G_r$ *and* $M_l = M_r$*,*
   a) $\frac{\partial b}{\partial M_l} = -\frac{\partial b}{\partial M_r}$, $\frac{\partial b}{\partial G_l} = \frac{\partial b}{\partial G_r}$, $\frac{\partial^2 b}{\partial M_l^2} = -\frac{\partial^2 b}{\partial M_r^2}$, $\frac{\partial^2 b}{\partial G_l^2} = -\frac{\partial^2 b}{\partial G_r^2}$*;*
   b) $\frac{\partial b}{\partial \beta} = 0$, $\frac{\partial^2 b}{\partial \beta \partial G_l} = -\frac{\partial^2 b}{\partial \beta \partial G_r} \leq 0$, $\frac{\partial^2 b}{\partial G_l^2} \leq \frac{\partial^2 b}{\partial G_l \partial G_r} \leq \frac{\partial^2 b}{\partial G_r^2}$*.*

According to Part 1 of Assumption 2, a regime expands territorially when it increases its military investment or moves its capital city toward its rival. Part 2 ensures that the SOCs are satisfied. Part 3 states that when the two regimes are symmetric, (a) similar changes made by each of them will generate similar effects; (b) the effect of an increase in β on the border is zero (since the effects on both sides cancel out each other); a higher β decreases the marginal gain of moving the capital closer to the competitor; the diminishing returns to moving the capital toward the competitor are strong enough so that the cross derivatives are always smaller in magnitude than the second derivatives.

The military also acts as a defense against the external threat by blocking it from spreading inland. Let $\kappa_i(x) = \lambda(\Lambda, x) - m(M_i, |G_i - x|, \beta)$ denote the **net damage** of the external threat at x (controlled by regime i). A location $x \in [0, 1]$ is **protected** by regime i from the external threat originating from 0 if there exists $0 \leq \hat{x} \leq x$ such that $\kappa_i\left(\hat{x}\right) \leq 0$. A location $x \in [0, 1]$ is **protected** by regime i from the external threat originating from 1 if there exists $x \leq \hat{x} \leq 1$ such that $\kappa_i\left(\hat{x}\right) \leq 0$. Let $\mathbb{D}_i = \{x \in [0, 1] : x \text{ is protected by regime } i\}$.

We assume that if less than δ fraction of the continent is protected, then a revolution occurs and all regimes in the continent receive negative payoffs.[25] To formalize the revolution constraint,

[25]This assumption, common in models of political economy, captures the idea that regimes that disregard the welfare of the population will be overthrown by a revolution, but revolutions involve overcoming collective-action

we assume that regime i's net revenue is $-\infty$ unless at least $\delta \in [0,1]$ fraction of the continent is protected. Under an empire, the net revenue of regime e is:

$$V_e = \begin{cases} y - c\left(M_e\right) & \text{if } |\mathbb{D}_e| \geq \delta, \\ -\infty & \text{otherwise,} \end{cases}$$

Under interstate competition, the net revenues of regimes l and r are given by:

$$V_l = \begin{cases} by - c(M_l) & \text{if } b > 0,\ |\mathbb{D}_l| + |\mathbb{D}_r| \geq \delta \text{ and } G_l \leq G_r, \\ -\infty & \text{otherwise,} \end{cases}$$

and

$$V_r = \begin{cases} (1-b)\, y - c(M_r) & \text{if } b < 1,\ |\mathbb{D}_l| + |\mathbb{D}_r| \geq \delta \text{ and } G_l \leq G_r, \\ -\infty & \text{otherwise,} \end{cases}$$

respectively.

3.2 Equilibrium

We now derive two propositions that characterize equilibrium outcomes when (1) $S = 1$; and (2) $S = 2$. The proofs of these propositions are provided in Appendix A.2.1–A.2.4.

First, consider the optimization problem under political centralization ($S = 1$). Regime e first decides the location of its capital $G_e \in [0,1]$ and then decides its military investment $M_e \geq 0$ to maximize net revenue V_e. Since this is a two-stage decision process, we employ backward induction to solve for the optimal solution.

Next, consider a two-stage game with interstate competition ($S = 2$). Regimes l and r simultaneously choose their capital cities $G_l \in [0,1]$ and $G_r \in [0,1]$. After observing the capital city locations, the regimes simultaneously make military investments $M_l \geq 0$ and $M_r \geq 0$. Again we employ subgame-perfect equilibrium as the solution concept.

Consider $x^*(\Lambda) \in [0,1]$ such that $\lambda(\Lambda, x^*(\Lambda)) = 0$. In other words, $x^*(\Lambda)$ is the leftmost location where the gross damage caused by the threat emanating from the left is zero. If such $x^*(\Lambda)$ does not

problems and therefore require support from a threshold population of $1-\delta$ to be successful (Tullock, 1971; Alesina and Spolaore, 2003; Acemoglu and Robinson, 2006). It is also consistent with the Confucian belief that the legitimacy of a government is contingent upon its ability to protect the people from harm and tax reasonably so that the people can maintain a constant means of livelihood. A government that loses this ability loses its 'mandate from heaven', and the people would therefore be entitled to depose it (Mencius, 2004).

exist, let $x^*(\Lambda) \equiv 1$. Consider Λ_I and $\bar{\Lambda}_I$ such that $x^*(\Lambda_I) = 1 - \delta/2$ and $x^*\left(\bar{\Lambda}_I\right) = 1 - \delta$ respectively. Let $\hat{\delta}$ denote the fraction of the continent that is protected from the external threat in equilibrium (i.e. $\hat{\delta} = |\mathbb{D}_e|$ when $s = 1$ and $\hat{\delta} = |\mathbb{D}_l| + |\mathbb{D}_r|$ when $s = 2$).

Proposition 1 (Empire)**.** *Under a two-sided threat of size Λ:*

1. *There exists Λ_I such that for all $\Lambda \leq \Lambda_I$, $M_e^* = 0$, $G_e^* \in [0, 1]$, and $\hat{\delta} \geq \delta$;*
2. *There exists $\Lambda_{II} > \Lambda_I$ such that for all $\Lambda_I < \Lambda \leq \Lambda_{II}$, $G_e^* = 1 - x^*(\Lambda) - \delta$, $M_e^* > 0$, and $\hat{\delta} = \delta$;*
3. *If $\Lambda > \Lambda_{II}$, $G_e^* = \frac{1}{2}$, $M_e^* > 0$, and $\hat{\delta} = \delta$;*

Under a one-sided threat of size Λ:

4. *There exists $\bar{\Lambda}_I$ such that for all $\Lambda \leq \bar{\Lambda}_I$, $M_e^* = 0$, $G_e^* \in [0, 1]$, and $\hat{\delta} \geq \delta$;*
5. *If $\Lambda > \bar{\Lambda}_I$, $G_e^* = 1 - \delta$, $M_e^* > 0$, and $\hat{\delta} = \delta$.*

Proposition 2 (Interstate Competition)**.** *Consider the symmetric equilibrium where $M_l^* = M_r^*$ and $G_l^* = 1 - G_r^*$. Under a two-sided [or one-sided] threat of size Λ:*

6. *There exists $\Lambda_{III}(\beta)$ [or $\bar{\Lambda}_{III}(\beta)$] such that if $\Lambda \leq \Lambda_{III}(\beta)$ [or $\Lambda \leq \bar{\Lambda}_{III}(\beta)$], the revolution constraint does not bind and $\hat{\delta} \geq \delta$. The equilibrium military investments and location of capitals are the same as in the case when $\Lambda = 0$;*
7. *If $\Lambda > \Lambda_{III}(\beta)$ [or $\Lambda > \bar{\Lambda}_{III}(\beta)$], the revolution constraint binds and $\hat{\delta} = \delta$.*

3.3 Implications for Political Centralization or Fragmentation

Proposition 1 implies that if the external threat is sufficiently weak (i.e., when $\Lambda \leq \Lambda_I$ as in Case 1 and $\Lambda \leq \bar{\Lambda}_I$ as in Case 4), the empire ignores it and makes zero military investment. This is because the sole motivation for the empire to invest in the military is to keep δ fraction of its population protected, so as to prevent a revolution. If the threat does not affect more than $1-\delta$ of the population, the empire has no incentive to make costly military investment. In all other situations (2, 3, and 5), the threat is **meaningful** and the empire builds a military and carefully chooses the location of its capital city to meet the threshold of protecting δ of the population.

By contrast, with or without the external threat, regimes in a competitive state system have to invest in the military to gain and maintain territory. Proposition 2 states that unless the external threat is sufficiently severe (Case 7), regimes l and r do not have to make additional military investments to protect their populations as their existing military capacity—built up as a result of competition among themselves—already meets this requirement.

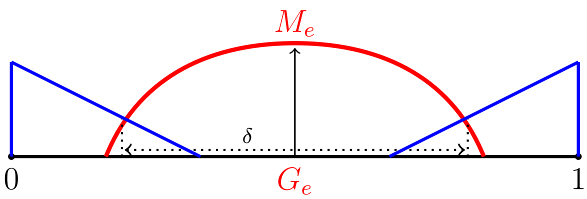


Figure 9: Optimal military investment under political centralization.

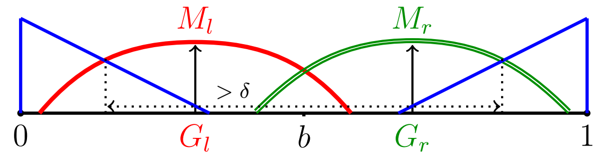


Figure 10: Optimal military investment under political fragmentation.

Put together, Propositions 1 and 2 indicate that political centralization and fragmentation have different strengths and weaknesses. First, in the absence of external threats, political fragmentation is wasteful from a static perspective and there are Pareto gains to be reaped if competitive regimes coordinate among themselves to reduce their military spending:

Implication 1 (**Wastefulness of interstate competition**). *If $\Lambda = 0$, military investment is zero under an empire but strictly positive under interstate competition.*

However if an external threat is present, an empire will spend a non-zero amount on the military only if the threat is sufficiently meaningful ($\Lambda > \Lambda_I$ if the threat is two-sided, or $\Lambda > \bar{\Lambda}_I$ if it is one-sided), and it will only provide protection to a fraction δ of the population so as to satisfy the revolution constraint (Figure 9). By contrast, in a competitive state system ($s = 2$), the competition-induced over-investment in the military may result in a larger-than-δ fraction of the continent being defended from external threats (Figure 10). Hence:

Implication 2 (**Robustness of interstate competition**). *If $\Lambda > 0$, interstate competition protects a weakly bigger interval of the continent than an empire does.*

Proposition 1 also suggests that the choice of an empire's capital city is influenced by the nature of the external threats that it confronts. In the absence of external threats, it does not matter where the empire's capital city is located. However, if the empire faces a meaningful one-sided threat, it will locate its capital city at $G_e^* = 1 - \delta$ to contain the threat. The higher is δ, the closer the capital city is to the frontier where the threat originates. Hence:

Implication 3 (**Locational choice of capital city**). *Under a one-sided external threat of size $\Lambda > \bar{\Lambda}_I$, it is not optimal for an empire to locate its capital city at the center, i.e., $G_e^* \neq 0.5$.*

Theoretical and empirical studies generally argue that capital cities should be centrally located to maximize tax revenue or improve governance (Alesina and Spolaore 2003; Olsson and Hansson 2011; Campante et al. 2014; Campante and Do 2014).[26] Implication 3 suggests that if the state is expected

[26] We can reproduce this result by introducing a cost of tax collection that is a weakly convex function of distance—in the absence of external threats, the empire will always locate its capital at the center of the continent, i.e., $G_e^* = 0.5$.

by its subjects to provide public goods, then, as long as the effectiveness of public goods provision differs with the location of provision, it may be optimal to separate the political center of the empire from its economic or population center.[27] In Section 5.1, we provide a historical discussion in light of this prediction.

3.4 *Stability of Political Centralization or Fragmentation under Different Threat Scenarios*

We define a regime as (fiscally) **viable** if its equilibrium net revenue is non-negative. When $S = 1$ and regime e is nonviable, political centralization is not a **stable** outcome. Even if an empire emerges, it cannot last. Likewise, when $S = 2$ and if one or both regimes are nonviable, political fragmentation will be unstable. Proposition 3 characterizes the fiscal viability of empires and competitive states under different threat scenarios:

Proposition 3. ***(Viability)***

1. *Under a one-sided threat,* $V_e^* > V_l^* + V_r^*$*;*
2. *Under a two-sided threat, when* $\Lambda \geq \Lambda_{II}$ *and* β *is sufficiently large,* $V_e^* < V_l^* + V_r^*$*.*

Proposition 3 gives rise to our main results on how the nature of a continent's external threat shapes the possibilities of political centralization and fragmentation:

Implication 4 (Stability under one-sided threat). *Under a one-sided threat, there exist threshold levels* $\dot{\Lambda}$ *and* $\ddot{\Lambda}$ *such that if* $\Lambda \leq \dot{\Lambda}$*, political centralization and political fragmentation are stable; if* $\dot{\Lambda} < \Lambda \leq \ddot{\Lambda}$*, political centralization is stable while political fragmentation is not; if* $\Lambda > \ddot{\Lambda}$*, political centralization and political fragmentation are unstable.*

Implication 5 (Stability under two-sided threat). *Under a two-sided threat, if* β *is sufficiently large, there exist threshold levels* $\dddot{\Lambda}$ *and* $\ddddot{\Lambda}$ *such that if* $\Lambda_{II} \leq \dddot{\Lambda} < \Lambda \leq \ddddot{\Lambda}$*, political fragmentation is stable while political centralization is not; if* $\Lambda > \ddddot{\Lambda}$*, political centralization and political fragmentation are unstable.*

Trivially, if an external threat is not meaningful—in that no military investment is required for the revolution constraint to be satisfied—then it has no impact on the viability of regimes. Whether the continent is politically centralized or fragmented does not matter. Both outcomes are stable.

However, if the threat is meaningful, military investment will be required to protect the population. As such, the more severe the threat, the less viable the regimes are. Proposition 3 implies that the nature of the threat (i.e., its direction and severity) helps define the continent's political possibilities—because it affects the viability of regimes under political centralization and fragmentation differently.

[27]This is true whether or not the cost of tax collection varies with distance. See Footnote 26.

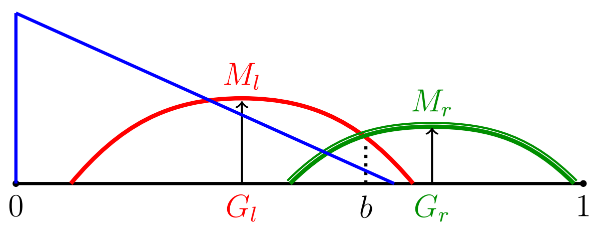


Figure 11: A one-sided, severe threat jeopardizes the viability of regime l, who has to compete with regime r for the continent's resources.

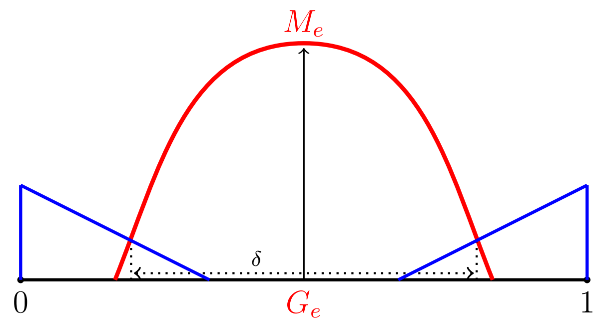


Figure 12: Even when threats are moderate, the empire has to invest heavily to secure two borders from one center if β is large.

Specifically, if the threat is one-sided, the stability of political centralization outlasts that of political fragmentation at high threat levels (Implication 4; see Figure 11); in the case of a two-sided threat, the reverse is true in the absence of modern technologies that facilitates the projection of military strength over long distances (Implication 5; see Figure 12).

Importantly, the two scenarios depicted above—a one-sided threat in the first and a two-sided threat in the second—are analogous to China's and Europe's external environments respectively.

3.5 *Other Results: Taxation*

We now endogenize taxation. Previously, the total amount of taxes paid in the continent was always equal to y. Suppose regime i has the option of reducing the tax burden of its people by $R_i \geq 0$. By keeping the population content, lowering taxes eases the revolution constraint. Consider an external threat emanating from 0 that is blocked by regime i at $x^{**} \in [0,1]$, i.e., $m(M_i, |G_i - x^{**}|, \beta) \geq \lambda(\Lambda, x^{**})$. The threat does not cause an individual at $x < x^{**}$ to engage in revolution as long as:[28]

$$\underbrace{R_i}_{\text{tax reimbursement}} + \underbrace{m(M_i, |G_i - x|, \beta)}_{\text{military protection}} - \underbrace{\lambda(\Lambda, x)}_{\text{damage from threat}} \geq 0.$$

When $S = 1$, as long as the external threat is meaningful so that the empire invests on the military, the revolution constraint always binds in equilibrium. We show in Appendix A.2.5 that if the cost function of military investment is sufficiently convex, the empire will provide some tax reimbursement instead of relying solely on building the military to satisfy the revolution constraint.

By contrast, when $S = 2$, the revolution constraint may not bind in equilibrium, in which case regimes l and r will set $R_l^* = R_r^* = 0$.

[28]We define the revolution constraint for an external threat emanating from 1 in a symmetric way.

Consider the two scenarios analogous to China and Europe: (1) the continent faces a one-sided, severe threat and is ruled by a single regime, and (2) the continent faces a two-sided, moderate threat and is ruled by two regimes. In the first scenario, the effective level of taxation will be $y - R_e$, where $R_e \geq 0$. In the second scenario, the level of taxation will remain at y. Hence:

Implication 6 (Taxation). *Taxation is weakly lower in a politically centralized continent confronting a one-sided threat than in a politically fragmented continent confronting a two-sided threat.*

Intuitively, an empire can secure the compliance of the population in two ways: it can protect them from the external threat or show restraint in fiscal extraction. If military investment is costly and lowering taxes offers a cheaper way to satisfy the revolution constraint, then it is in the empire's interest to underprovide military defense for territories that are distant and hence costly to defend. By contrast, under political fragmentation regimes invest in the military not only to protect their people from external threats, but also to maintain and extend their territory vis-à-vis other regimes. If interstate competition is intense, they need to invest heavily in the military. Lowering taxes will not reduce this need and as such they have no incentive to do so.

3.6 Other Results: Population Dynamics and Long-run Growth

Until now, we have assumed that external threats are always present. Suppose instead that the external threat is realized with some positive probability. Suppose also that each individual inelastically supplies labor to produce $y + \underline{y}$, where y is taxable and $\underline{y}$ is not. For individual x under regime i, the disposable income is $\bar{y} = \underline{y} + R_i - \kappa_i(x)$ where R_i is the tax reimbursed by regime i and $\kappa_i(x)$ is the net damage caused by the stochastic shock. Each individual chooses between private consumption c and producing n offspring to maximize her utility $u(c, n)$ subject to the budget constraint $\rho n + c \leq \bar{y}$, where ρ represents the cost of raising a child. We assume that c and n are complements and u is increasing and concave in both arguments. Standard optimization implies that the optimal number of children is $n = g(\bar{y})$ where $g' > 0$.

Of interest is the continent's population, which is given by:

$$N = \int_0^1 n dx = \int_0^1 g(\bar{y})\ dx\,.$$

Let N_E and N_F denote population in continents E and F respectively. The two continents are identical except that continent E is ruled by an empire ($S = 1$) and faces a one-sided threat of size Λ_E, while continent F is politically fragmented ($S = 2$) and faces a two-sided threat of size Λ_F.

When the external threat is not realized, the populations in the two continents are determined by $N_E = g(\underline{y} + R_e)$ and $N_F = g(\underline{y})$ respectively. Since $N_E > N_F$, population grows faster under the empire.

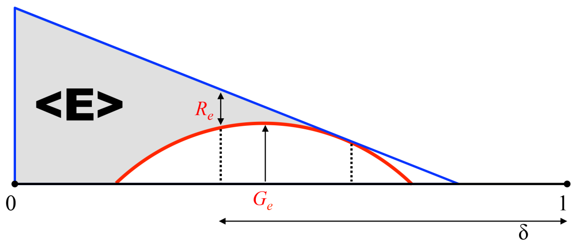


Figure 13: Under political centralization, there is a positive level of tax reduction, i.e. $R_e^* > 0$. Under political fragmentation, tax reduction is zero.

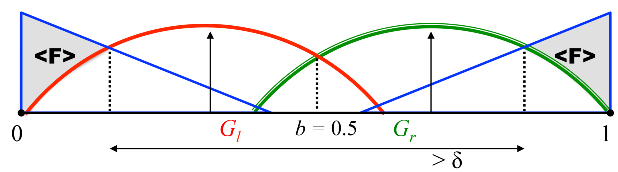


Figure 14: Under political fragmentation, the fraction of the population protected from invasion is at least δ. Under political centralization, it is at most δ.

However, the converse may be true if the external threat is realized. In this case, realized populations under empire and political fragmentation are given by, respectively:

$$N_E = \int_0^1 g(\underline{y} + R_e - \mathbb{1}_{x \notin \mathbb{D}_e}[\lambda(\Lambda_E, x) - m(M_e, |G_e - x|, \beta)])\ dx;$$

$$N_F = \int_0^1 g(\underline{y} - \mathbb{1}_{x<b, x \notin \mathbb{D}_l}[\lambda(\Lambda_F, x) - m(M_l, |G_l - x|, \beta)] \\ - \mathbb{1}_{x>b, x \notin \mathbb{D}_r}[\lambda(\Lambda_F, 1-x) - m(M_r, |G_r - (1-x)|, \beta)])\ dx.$$

Now $N_E < N_F$ if Λ_E is sufficiently large with respect to Λ_F.

For the purpose of illustration, let $u(c, n) = c^{1-\gamma} n^{\gamma}$. It can be shown that when the shock is realized, the populations under empire and under fragmentation are given by:

$$N_E = \frac{\gamma}{\rho} \cdot \left\{ (\underline{y} + R_e) - \underbrace{\int_{x \notin \mathbb{D}_e} \lambda(\Lambda_E, x) - m(M_e, |G_e - x|, \beta)\ dx}_{\text{Area}\langle E \rangle} \right\};$$

$$N_F = \frac{\gamma}{\rho} \cdot \left\{ \underline{y} - \underbrace{2 \cdot \int_{x<b, x \notin \mathbb{D}_l} \lambda(\Lambda_F, x) - m(M_l, |G_l - x|, \beta)\ dx}_{\text{Area}\langle F \rangle} \right\},$$

where Area $\langle E \rangle$ and Area $\langle F \rangle$ are illustrated in Figures 13 and 14.

If $\Lambda_E > \Lambda_F$, Area $\langle E \rangle$ is likely to be larger than Area $\langle F \rangle$ not only because the empire in continent E confronts a more severe external threat than the competing regimes in continent F, but also because the empire offers protection to only δ fraction of its population (and less than δ if tax reduction is offered), while the fraction of continent F that is protected is always weakly larger than δ (Implication 2). If Area $\langle F \rangle <$ Area $\langle E \rangle - R_E$, it follows that $N_E < N_F$.

Hence, comparing the two scenarios analogous to China and Europe, if the external threat is not realized, population grows faster under political centralization, but if the external threat is realized, a population contraction is also more likely under political centralization:

Implication 7 (**Population Change**). *Population change displays a higher variance in a politically centralized continent confronting a severe, one-sided threat than in a politically fragmented continent confronting a moderate, two sided threat.*

In interpreting our model, we have focused on external invasions. More generally, however, negative shocks could also stem from unforeseen political collapses and peasant rebellions in addition to invasions from outside. The central point we emphasize is that interstate competition results in a greater proportion of territory being protected than is the case under political centralization.

4 External Threats and Political Unification or Fragmentation: Empirical Evidence

The model predicts that the stability of political centralization and fragmentation is unaffected by external threats when these threats are sufficiently small; an environment with external threats originating from multiple fronts favors interstate competition; a unidirectional threat promotes political unification if the threat is severe. To test these predictions, we first investigate the empirical relationship between the frequency of nomadic attacks and political unification in China. Subsequently, we examine historical evidence from Europe for consistency with our predictions.

4.1 Empirical Evidence from China

To test our hypothesis that a severe and unidirectional threat from the steppe provided a recurring impetus for unification in China we exploit time series variation in political unification and fragmentation in Chinese history. We show that periods of more conflict with steppe nomads were positively associated with periods of political unification within China, and vice versa.

Data Sources and Definition of Variables In a recent paper, Bai and Kung (2011) show that nomadic incursions into China were correlated with exogenous variations in rainfall as climatic disasters such as droughts often triggered subsistence crises that drove the inhabitants of the ecologically fragile steppe to invade their settled neighbors for food. We make use of their data set and empirical strategy to test if there was a relationship between the frequency of nomadic attacks and political unification in China. This helps to ensure that our empirical evidence is robust and is not selectively adopted to suit our purpose.

Bai and Kung's data span 2,060 years (from 220 BC to AD 1839) and are drawn from four sources: *A Chronology of Warfare in Dynastic China* (China's Military History Editorial Committee, 2003), *A*

Table 1: List of Variables and Summary Statistics

Variable		Description	mean	s.d.
Unification	y_t	=1 if only 1 regime ruled China in decade t	0.59	0.49
#Regimes	y_t^{alt}	Average number of regimes in China proper in decade t (*log*)	0.39	0.54
Nomadic attacks	x_t	Frequency of attacks initiated by the nomads in decade t	2.53	3.50
Lower precipitation	z_{1t}	Share of years with records of drought disasters on the Central Plain in decade t	0.50	0.30
Higher precipitation	z_{2t}	Share of years with records of Yellow River levee breaches in decade t	0.18	0.21
Snow disasters	w_{1t}	Share of years with records of heavy snow on the Central Plain in decade t	0.12	0.14
Low temperature disasters	w_{2t}	Share of years with records of low-temperature calamities (e.g., frost) on the Central Plain in decade t	0.16	0.19
Temperature	w_{3t}	Average temperature in decade t	9.46	0.89
Nomadic conquest 1	w_{4t}	=1 if the Central Plain was governed by the nomads (317–589)	0.13	0.33
Nomadic conquest 2	w_{5t}	=1 if the Central Plain was governed by the nomads (1126–1368)	0.12	0.32
Nomadic conquest 3	w_{6t}	=1 if the Central Plain was governed by the nomads (1644–1839)	0.10	0.29
Time trend	w_{7t}	Decade: -22–183 (219 BC–1839)	80.5	59.6

Sources: Bai and Kung (2011) and Wei (2011).

Compendium of Historical Materials on Natural Disasters in Chinese Agriculture (Zhang et al., 1994), *A Concise Narrative of Irrigation History of the Yellow River* (Editorial Committee of Irrigation History of the Yellow River, 1982), and the *Handbook of the Annals of China's Dynasties* (Gu, 1995). Of these sources, the first three have been widely used in related research and are considered reliable sources while the fourth contains general historical information that can be easily verified.[29]

As listed in Table 1, the decadal variables Bai and Kung constructed include: (i) the frequency of nomadic attacks on China's Central Plain (Henan, Shanxi, Shaanxi, Hebei); (ii) two precipitation variables that measure the extent of severe droughts and floods in the Central Plain; (iii) other climatic control variables (snow and other low temperature disasters, temperature); (iv) a time trend. Drawing from Wei (2011), we add a new dependent variable: *Unification*, that takes the value of 1 if China was politically unified in a given decade and 0 otherwise.

ADL Estimation As a baseline to investigate the effects of nomadic invasions on political unification in China, we adopt a simple autoregressive distributed lag (ADL) model:

$$y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=0}^{q} \mu_i x_{t-i} + \epsilon_t \ , \tag{1}$$

where y_t is the dummy variable *Unification* and x_t is the number of nomadic incursions into China in decade t.

The ADL model is appropriate for our purpose because of its flexibility. Furthermore, it generates unbiased long run estimates and valid t-statistics even in the presence of endogeneity (Harris and Sollis, 2003). To validate our use of the ADL methodology, we use the Augmented Dickey-Fuller test to ensure that all variables are stationary. To determine the appropriate number of lags, we follow

[29] Bai and Kung (2011, Table A.2) provides a detailed check on the accuracy of the data on Sino-nomadic conflicts.

the general-to-specific approach proposed by Ng and Perron (1995) to seek the values of p and q in Equation 1 that minimize the Akaike Information Criterion (AIC), which occurs at $p = 3$ and $q = 1$.[30]

According to Implication 4, an increase in the severity of nomadic threat favors political unification. While this effect may not be immediate—in the short run, a spike in nomadic attacks could lead to the collapse of the central authority and the emergence of a host of succession states to fill up the political vacuum—Implication 4 suggests that increased nomadic attacks should have a long run positive effect on political unification. In other words, we expect $\mu_0 + \mu_1 + \mu_2 + ... + \mu_q > 0$.

In the ADL model reported in column (a) of Table 2, we find that the nomadic invasion variable and its lagged value are both statistically significant, but they carry opposite signs: an additional nomadic attack in decade t is associated with a 1.2% decrease in the probability of politically unification in China in the same decade, but an attack in the previous decade (at $t - 1$) is associated with a larger 1.96% increase in the probability of politically unification in decade t. Their joint F statistic is 5.32, which implies that one can reject the null hypothesis that the two coefficients are jointly zero at 1% significance level. In line with Implication 4, the relationship between nomadic invasions and political unification is positive in the long run: each additional nomadic attack is associated with an increase in the probability of political unification of 6.3% ($= \frac{-0.012+0.0196}{1-0.906+0.283-0.256}$). Given that China experienced an average of 2.5 nomadic attacks per decade, the relationship between nomadic invasions and its political unity is a meaningful one.[31]

In column (b) of Table 2, we deviate from the classic ADL model and introduce the control variables as used in Bai and Kung (2011) into our estimation equation, which now becomes:

$$y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=0}^{q} \mu_i x_{t-i} + \sum_{i=0}^{q} \psi_i^1 z_{1t-i} + \sum_{i=0}^{q} \psi_i^2 z_{2t-i} + \pi W_t + \epsilon_t \,, \tag{2}$$

where z_{1t} and z_{2t} are rainfall variables measuring droughts and floods and W_t is a vector of seven other climatic and historical control variables (see Table 1 for details). Since Bai and Kung (2011) detect a strong and robust relationship between the frequency of nomadic invasions and rainfall factors, the inclusion of rainfall variables in Equation 2 leads to multicollinearity and increases the standard errors of the estimates. However, we obtain very similar and statistically significant coefficient estimates.

VAR Estimation The estimations above establish correlation, not causation. In particular, one would suspect that political fragmentation could leave China divided and weakened, and therefore increase the likelihood of nomadic attacks. This source of endogeneity does not pose a problem for our argument—if indeed political fragmentation had the effect of increasing nomadic attacks,

[30]When implementing the general-to-specific approach, we choose $p = q = 10$ as the cut-off and check every combination of $p \leq 10$ and $q \leq 10$.

[31]The Durbin's h-test indicates that the errors are serially independent. In addition, the roots of the characteristic equation are all smaller than 1 and therefore the estimation model is 'dynamically stable'.

Table 2: ADL Model

	(a)	(b)
Dependent variable:	Unification	Unification
Unification: Lag 1	0.906***	0.875***
	(0.0651)	(0.0668)
Unification: Lag 2	-0.283***	-0.277***
	(0.0843)	(0.0837)
Unification: Lag 3	0.256***	0.202***
	(0.0630)	(0.0648)
Nomadic attacks	-0.0120**	-0.0108*
	(0.00605)	(0.00628)
Nomadic attacks: Lag 1	0.0196***	0.0182***
	(0.00604)	(0.00621)
Additional controls	No	Yes
Observations	203	203
R-squared	0.743	0.765
AIC	0.122	0.141

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table 3: VAR Model

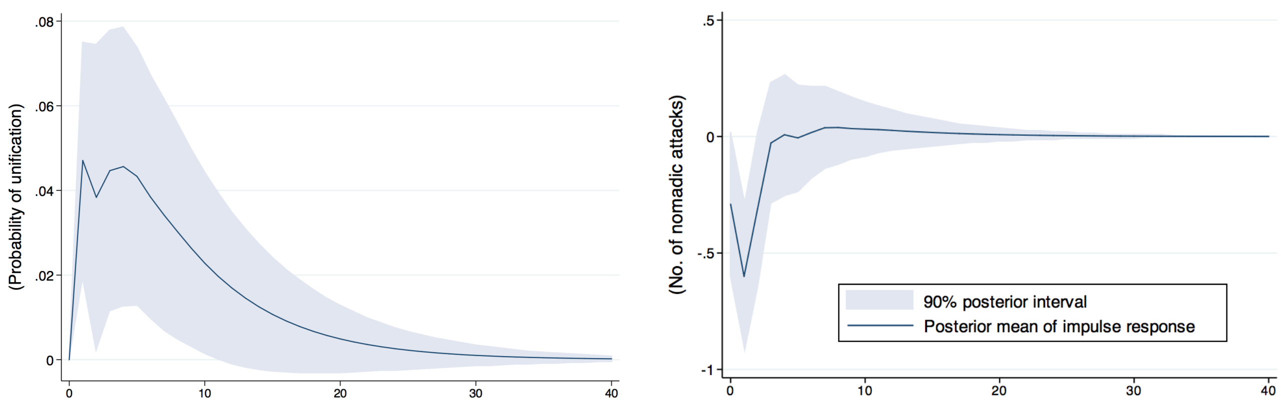
	(a)	(b)
Dependent variable:	Unification	Nomadic attacks
Unification: Lag 1	0.893***	-2.075***
	(0.0665)	(0.733)
Unification: Lag 2	-0.317***	1.631*
	(0.0848)	(0.935)
Unification: Lag 3	0.225***	0.377
	(0.0656)	(0.723)
Nomadic attacks: Lag 1	0.0176***	0.321***
	(0.00626)	(0.0690)
Nomadic attacks: Lag 2	-0.00701	0.257***
	(0.00657)	(0.0724)
Nomadic attacks: Lag 3	0.00602	-0.0108
	(0.00638)	(0.0703)
Additional controls	Yes	Yes
Observations	203	203

the true effect of nomadic invasions on China's political unity would be even larger. Nonetheless, to investigate further, we implement the following vector autoregression (VAR), which models the simultaneity of our dependent and main explanatory variables explicitly:

$$\begin{aligned}\begin{bmatrix} y_t \\ x_t \end{bmatrix} &= \begin{bmatrix} \phi_0 \\ \mu_0 \end{bmatrix} + \begin{pmatrix} \phi_1^1 & \mu_1^1 \\ \phi_1^2 & \mu_1^2 \end{pmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{pmatrix} \phi_2^1 & \mu_2^1 \\ \phi_2^2 & \mu_2^2 \end{pmatrix} \begin{bmatrix} y_{t-2} \\ x_{t-2} \end{bmatrix} + \\ &+ \begin{pmatrix} \phi_3^1 & \mu_3^1 \\ \phi_3^2 & \mu_3^2 \end{pmatrix} \begin{bmatrix} y_{t-3} \\ x_{t-3} \end{bmatrix} + \begin{pmatrix} \psi_0^1 & \psi_0^2 \\ \psi_0^1 & \psi_0^2 \end{pmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} + \Pi W_t + \begin{bmatrix} \epsilon_{t-1} \\ \epsilon_{t-2} \end{bmatrix} . \end{aligned} \tag{3}$$

As with the previous estimations, we select the lagged values that minimize the AIC. We also checked for autocorrelation and that the eigenvalues lie inside the unit circle (hence the VAR model is 'dynamically stable'). As Table 3 illustrates, the estimates from the VAR model share the same order of magnitude with the results from the ADL estimations. The coefficient estimate of Lag-1 nomadic attack is 0.0176 in column (a) of Table 3, compared with 0.0196 and 0.182 in columns (a) and (b) of Table 2. The Wald test statistic of the coefficients on the lags of nomadic attacks is 332.6. Hence, we reject the null hypothesis that nomadic attacks did not Granger-cause political unification at 1% significance level.

As a further check on the interaction between nomadic attacks and political unification, we estimate their impulse response functions. Figure 15 reports the posterior means and the 90% posterior intervals for horizons of 40 decades. As Figure 15a illustrates, an exogenous, one-standard-deviation increase in nomadic attacks increases the probability of political unification in China in a persistent manner. The estimated effect remains statistically significant for 10 decades after the



(a) Impulse response of unification to nomadic attacks (b) Impulse response of nomadic attacks to unification

Figure 15: Estimated impulse responses

shock, and dies out after 20 decades.

Meanwhile, Figure 15b suggests that political unification decreases the frequency of nomadic attacks on China. However, the estimated effect is limited and is statistically significant only for one period (at $t+1$). Hence, the relationship between political unification and nomadic attacks was an asymmetric one: the impact of nomadic attacks on political unification was more salient than the reverse. This is consistent with other empirical studies, which find that nomadic invasions on China were often climate-induced (Bai and Kung, 2011; Zhang et al., 2015). Importantly, if political unification did have an effect on the number of nomadic attacks, it would lead to fewer nomadic attacks and not more—incorporating this effect would only strengthen our argument that the nomadic threat provides an impetus that helped promote political unification in China.

In Appendix A.6, we test the robustness of the above results by using the *log* number of regimes in China proper as an alternative dependent variable. Consistent with the above results, an increase in nomadic attacks is associated with a decrease in the number of regimes in China proper.

4.2 Historical Evidence from Europe

We are unable to replicate the above empirical exercise for Europe because data on the number of regimes in Europe only exists on a per century basis. However, European historical patterns do conform to predictions of our theoretical model.

Europe has historically been politically fragmented; the closest Europe came to be ruled by a unified political system was under the Roman Empire. The rise of Rome parallels the rise of the first empire in China (Scheidel, 2009). In terms of the model, one advantage Rome had over its rivals in the Hellenistic world was a relatively less convex cost function of military investment—Rome's ability to project power and increase its resources of manpower was unequaled among European states in antiquity (Eckstein, 2011). Thus, Rome was able to impose centralized rule upon much of Europe. Our model suggests that two factors can account for the decline of Rome: (1) over time, Rome's military advantage declined relative to the military capacities of its rivals such as the Persian empire or the Germanic confederacies; and (2) these rising threats came from multiple directions along Rome's long border.[32] Like episodes of dynastic and imperial collapse in China, the fall of the western Empire was associated with political disintegration and economic collapse across Europe (Ward-Perkins, 2005). However, unlike in China, all subsequent attempts to rebuild the Roman Empire failed; our model sheds light on why this was the case.

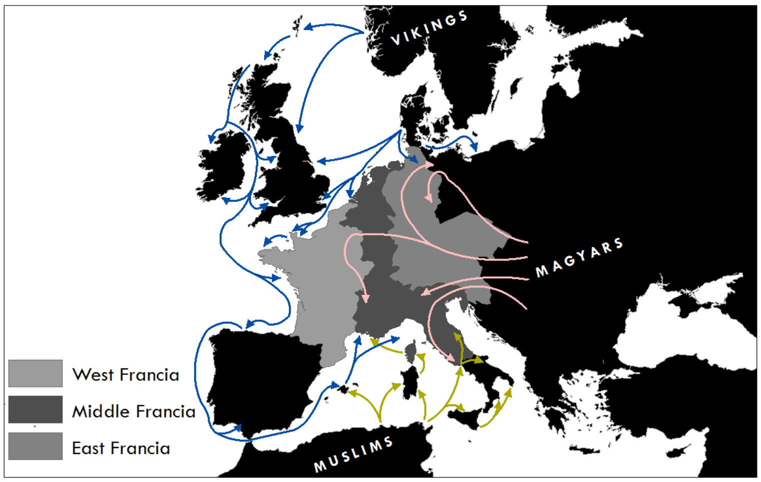


Figure 16: Viking, Magyar, and Muslim Invasions of Western Europe in the Ninth and Tenth Centuries; The Carolingian Empire after the partition of AD 843.

The most successful subsequent attempt to build a European-wide polity was the creation of a Frankish empire by the Carolingians. The Carolingian dynasty was established by Pippin III (r. 752–768) and under the reign of Charlemagne (r. 768-814) came to control an empire that spanned France, parts of Spain and much of Italy and central Europe (Collins, 1998; McKitterick, 2008).

Consistent with our model, the Carolingian empire was not long-lasting. It went into decline as the successors of Charlemagne struggled to deal with the external threats posed by the Magyars, the Vikings, and the Muslims from different fronts (Morrissey, 1997).[33] The empire was divided in 843 (Figure 16). In East Francia, a different dynasty, the Ottonians, came to power as a response to the repeated Magyar invasions and established the Holy Roman Empire. At its height in the eleventh century, it comprised modern day Austria, Czech Republic, Germany, Italy, the Low Countries, and Switzerland. Increasingly, emperors based in Germany found it difficult to control their Italian

[32]These claims are consistent with the vast historical scholarship on this topic (see Heather, 2006).

[33]As Spruyt observes 'the breakup of central authority coincided with the increasing raids by Magyars, Saracens, and Vikings' (Spruyt, 1994, 37). Viking raids began in the reign of Charlemagne but greatly increased in ambition after the death of Louis the Pious exposed the fragility of the Carolingian hegemony and hastened the emergence of localized regional power centers.

provinces and by the thirteenth century, the Holy Roman Empire was no more than a loose federation of German principalities.

Incidentally, the threats posed to the Europeans by the Vikings and the Muslims also receded after the eleventh century. One could argue that from then on, Western Europe no longer experienced meaningful multi-sided external threats. If this interpretation is correct, our model predicts that the status quo of political fragmentation would persist, and it did. The Mongol invasion of Europe in the thirteenth century was too brief to provide a sustained impetus toward European unification. However, the less dramatic but more sustained rise of the Ottoman empire after the fifteenth century serves as yet another test of our model and it provides further supporting evidence that our mechanisms are relevant. Iyigun (2008) shows that the external threat of invasion from the Ottomans between 1410 and 1700 reduced the frequency of interstate warfare in Eastern Europe. Indeed, a comparison of the political maps of Central and Eastern Europe of the fourteenth century and seventeenth century indicates that 'a significant degree of political consolidation accompanied the Ottoman expansion in continental Europe' (Iyigun, 2008, 1470).

5 Applying the Model

We are now in a position to use our theory to offer new insights into the location choice of capital city, differential levels of taxation, and differential patterns of population change at the two ends of Eurasia.

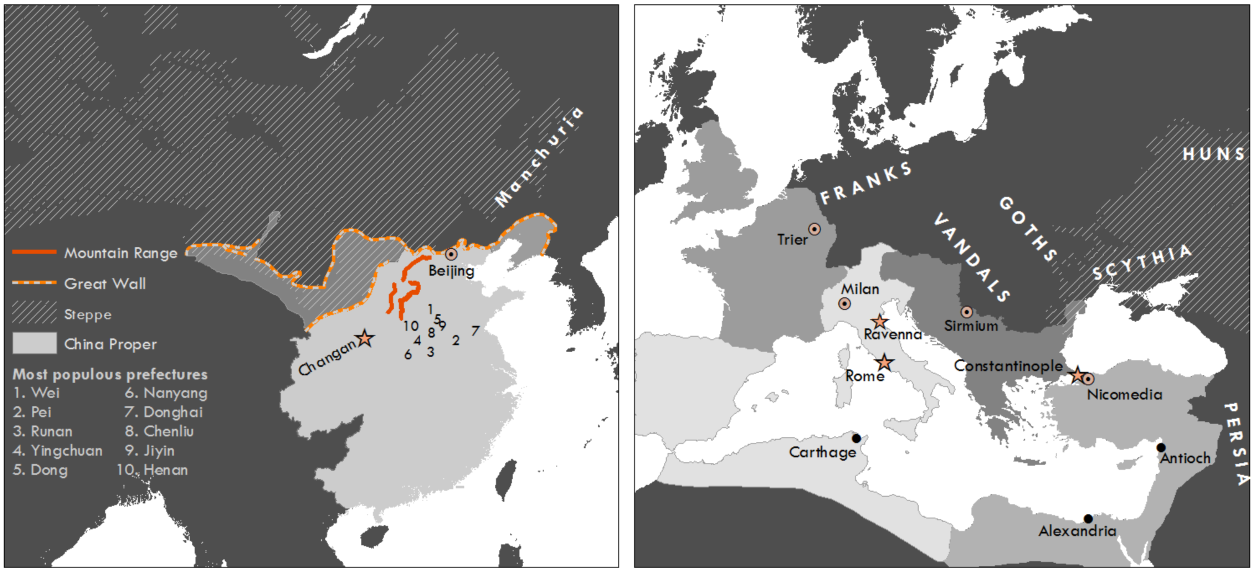
5.1 Locations of Capital Cities

Our model predicts that the threat of external invasion is an important determinant of the location of an empire's capital. There are numerous examples of empires changing capitals to confront their external enemies more effectively; we focus on examples from China and Europe.

Consistent with our model, for most of its history, China's capital city was located in its northern or northwestern frontier instead of the populous Central Plain or Lower Yangzi Delta. For the 1,418 years between 221 BC and AD 1911 when China was under unified rule, Beijing and Changan (modern day Xi-an) served as its national capital for 634 years and 553 years respectively, or together 8.4 years out of every 10 years (Wilkinson, 2012).

Changan was China's preeminent political center in the first millennium. It was the capital city of the unified dynasties of Qin (221–206 BC), Former Han and Xin (202 BC–AD 23), Sui (581–618), and Tang (618–907). Figure 17a illustrates two salient characteristics of Changan's geographical location that buttress our argument: (1) it was not the population or economic center of the empire;[34] and (2)

[34]During the Han dynasty, for example, only 3 million people, or around 4 percent of the Chinese population,



(a) Han Dynasty (206 BC–220 AD)

(b) Roman Empire, Tetrarchy, circa 300 AD

Figure 17: Capital cities in China and the Roman Empire. Panel (a) depicts the Han dynasty's capital city, Changan, and its most populous prefectures. Beijing replaced Changan as China's preeminent political center in the second millennium. Panel (b) depicts the major cities of the Roman Empire. During the Tetrarchy period, there were four capitals: Trier, Milan, Sirmium, and Nicomedia. Constantinople and Ravenna were the capitals of the Eastern and Western Empires respectively. Carthage, Alexandria, and Antioch were the largest cities after Rome itself. The maps are adapted from Herrmann (1966), Skinner et al. (2007), and Talbert (2000).

it shielded China's populous Central Plain from nomadic invasions by virtue of its strategic location between the steppe and central-eastern China.

In the second millennium when China's threat from the north shifted from Inner Asia eastward to the semi-nomadic lands of Manchuria, Beijing replaced Changan as the new political center of China. The emergence of Beijing was due to its proximity to the northern frontier—the Chinese thinker and political theorist Huang Zongxi (1610–1695) likened making Beijing the national capital to having the emperor guard his empire's gates (Huang, 1993).

For the European case, our evidence comes from the Roman Empire—the single long-lasting empire to span much of the continent in European history. We find strong support for this prediction of the model. The Roman Republic and Empire expanded symmetrically from the city of Rome over several centuries to encompass the entirety of the Mediterranean and western Europe. Over time therefore, the location of the capital became less and less convenient from the point of view of military operations. This was not a major issue during the first century of the Empire, but as the severity of the external threats facing the empire grew from the mid-second century onward emperors

resided in Guanzhong, the region where Changan was located. By contrast, the Guandong region in central-eastern China was home to 60 percent of the empire's subjects.

spent less and less time in Rome and they eventually set up other capital cities in which to reside.[35] Figure 17b depicts both the old capital of Rome and the capitals established during the latter part of the third century: Trier on the Rhine frontier, Milan at the entrance of the Alps shielding Italy from invasion, Sirmium on the Danube frontier, and Nicomedia in the east. Consistent with the predictions of the model, therefore, it was the threat of invasion that drove the choice of capital location for the Roman Empire.[36]

Importantly, the choice of these new capital cities did not correspond to the largest cities. After Rome, which remained the biggest city in the empire through this period,[37] the most populous cities in the empire were Alexandria with around 600,000 inhabitants, and Antioch and Carthage with between 300,000–500,000 people (Scheidel, 2013, 78). However, with the exception of Antioch these cities were far from the frontiers and were not chosen as capital cities for this reason.[38]

When the emperor Constantine (r. 306–338) established a new permanent capital at the small Greek city of Byzantine, renamed as Constantinople, he chose this location not because it had any economic significance at the time, but because it was close to both the eastern frontier of the empire and to the important Danube front where the empire faced some of its most determined enemies.[39]

Interestingly, the lessons of Chinese and Roman history also provide ample evidence to support the assumption in our model that rules out the possibility of a state maintaining two or more comparable political-military centers. During the mid-Tang dynasty, the Xuanzong emperor (r. 712–756) implemented a polycentric political-military system and devolved much of the central government's political authority to frontier military governors with the goal of improving military responsiveness and effectiveness. However, Xuanzong's favorite and most powerful frontier governor, An Lushan, infamously revolted in 755 as the military might of An's army fed popular suspicion of his political ambitions, which ironically compelled An to revolt. A similar development took place during the early Ming dynasty with the implementation of a de facto twin-capital system in which the emperor resided in the populous south and his uncle, the Prince of Yan, coordinated border defense in the strategic north. The arrangement again proved unstable and mutual suspicion led to the outbreak of a bloody civil war in 1399 with the Prince of Yan emerging as the eventual victor. To prevent history from repeating itself, the usurper moved the capital city from Nanjing (the 'southern

[35] For example, Gallenius (r. 253–268) did not visit Rome until the fifth year of his reign while Diocletian (r. 284–306), who established his capital at Nicomedia in modern day Turkey, did not visit Rome until the twentieth year of his reign (Rees 2004, 28 and Goldsworthy 2009, 162).

[36] Rees (2004, 27) notes that '[t]he motivation for this move away from the ancient capital seems to have been strategic. . . the north Italian cities, Milan, Ravenna, and Aquileia, were closer to the main military areas than Rome was.'

[37] Though its population declined from a peak of 1 million in the 2nd century to around 750,000 in the 4th century.

[38] Antioch was in fact a major base to fight the Persians even if it was not officially a capital city. For example, the Emperor Julian (r. 361–363) spent nine months in Antioch preparing for his invasion of Persia.

[39] For example, see Odahl (2004, 232) and Goldsworthy (2009, 186).

capital') to Beijing (the 'northern capital') in order to maintain direct control of the large army along the northern border.

Similarly, the Roman Empire was not able to maintain multiple capitals within a single empire for a long period of time. The fourfold division of the empire inaugurated by Diocletian (r. 284–306), known as the Tetrarchy, did not last long beyond his retirement in 306 and led to a series of civil wars that only ended with Constantine's reunification in 324. Civil wars reoccurred during periods of imperial division in 337–350, 360–361, 383–388, 392–395. By the end of the fourth century, the centrifugal forces affecting the empire led to the permanent division of the empire into East and West and from this point on, the two empires coexisted as separate political entities until the fall of the Western Empire in 476.

5.2 Taxation

Our model predicts that taxation would be higher in Europe relative to China (Implication 6). This contradicts traditional comparative accounts of Europe and China, which complained that economic development in China was retarded by high taxes (e.g. Jones 2003), but it is consistent with recent scholarship in economic history. Tax rates in Europe were high and especially so in the Dutch Republic and England after 1689 (Hoffman and Norberg, 1994; Bonney, 1999).[40] By contrast, taxes were comparatively low in China. Karaman and Pamuk (2013) provide data on tax revenues for a range of European countries. Table 4 depicts this data in conjunction with estimates of per capita and total tax revenue from China (Sng, 2014). Tax revenue per capita in France was lower than in either the Dutch Republic or England, but it remained much higher than in China.[41] The average European per capita level of taxation as measured in silver was roughly four times higher compared to China. As China was a net importer of silver, the value of silver in China might have been higher than in Europe. Following Ma (2013), we use the bare-bones subsistence basket constructed in Allen et al. (2011) to estimate the tax burden in Europe and China and obtain similar results. Clearly, as Implication 6 suggests, taxation was lighter under politically centralized China than it was in fragmented Europe.[42]

[40]This increase in tax revenue obviously implies that the tax revenues collected by the central state were lower before 1700. However, this does not mean that the total level of taxation was lower when one includes feudal dues, local taxes, and tithes to the Church. Much of the increase in tax revenue collected by centralized states came at the expense of these forms of taxation that were generally restricted and abolished in the seventeenth and eighteenth centuries. For a comparison of taxes in England and France in the seventeenth century see Johnson and Koyama (2014a).

[41]Johnson and Koyama (2014b) document the increase in tax revenues at a regional level in France throughout the seventeenth century.

[42]Our hypothesis complements a strand of the literature that attributes light taxation in China to a severe principal-agent problem in its government (Kiser and Tong, 1992; Ma, 2013; Sng, 2014; Sng and Moriguchi, 2014). Instead of focusing on agency costs in tax collection as this literature does, we look at agency costs in public goods provision (as argued in Section 3.1, agency costs in military control lead to the deterioration of military strength over distance) and

	Per Capita Tax Revenue (silver grams)			Total Tax Revenue (silver tons)		
	1700	1750	1780	1700	1750	1780
England	91.9	109.1	172.3	559.4	821.1	1627.3
France	43.5	48.7	77.6	878.2	1081.2	1962
Dutch Republic	210.6	189.4	228.2	400.6	367.6	466.8
Spain	28.6	46.2	59.0	219.2	439.3	642.5
European average	52.1	58.0 (27%)	77.3	278.2	403.2	711.5
China	10.4	11.8 (6%)	9.2	1812.1	2633.3	2769.3

Table 4: Per capita tax revenue in grams of silver. European average tax revenue includes Venice, Austria, Russia, Prussia, and Poland-Lithuania in addition to England, France, Dutch Republic, and Spain. Sources: Karaman and Pamuk (2013) and Sng (2014). In parentheses we include a comparison of per capita tax revenue as a proportion of 'bare-bones' subsistence in 1750 as measured by Allen et al. (2011).

By and large, the taxes raised by European states were spent on warfare.[43] Scholars have long noted that Europe was the 'seat of Mars' (Tilly, 1990; Voigtländer and Voth, 2013). In China as in Europe, the majority of government revenue was spent on the military (Sng, 2014, Table 5). The main difference was that China's total military spending was much lower. As Vries (2012) observed, 'with roughly twenty times as many inhabitants, China, in real terms, per year on average only spent roughly 1.8 times as much on the military as Britain did during the period from the 1760s to the 1820s. Per capita in real terms Britain thus spent more than ten times as much on its army and navy than China' (Vries, 2012, 12).

5.3 Population cycles in China and Europe.

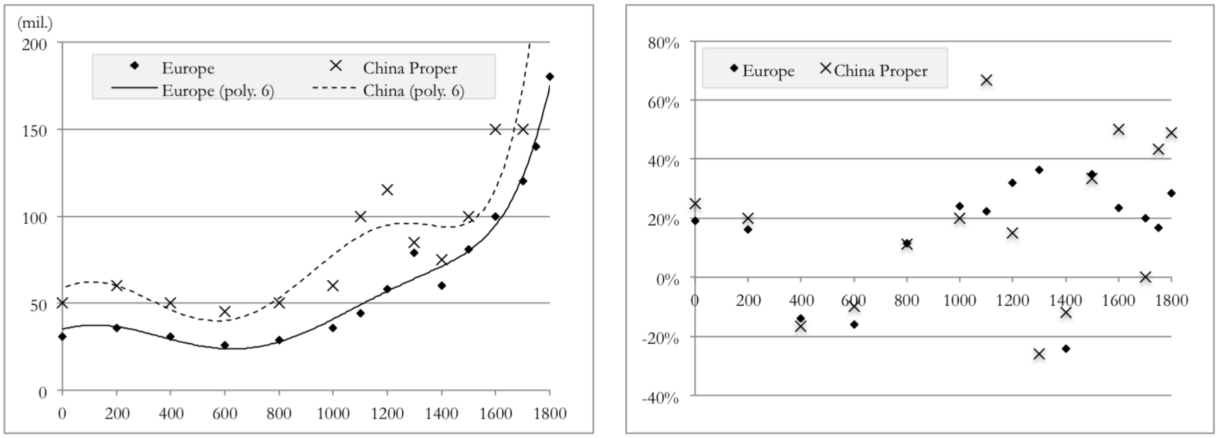
Our model also offers new insights into the consequences of the persistence of political centralization in China and fragmentation in Europe. Implication 7 predicts that population growth should be more variable under political centralization because political centralization is associated with lower taxes during peacetime but also greater vulnerability to external shocks. We provide evidence in support of this proposition by drawing on population data from China and Europe.

Pre-modern population data is of variable quality. McEvedy and Jones (1978) provide imperfect but comparable population estimates for both China and Europe for the past two thousand years. Figure 18a presents these estimates. It shows that the population growth of China was more variable than that of Europe. Figure 18b, which shows the percentage population change, confirms this finding. It is evident that the time series of Chinese population display greater variance. Interestingly, there

derive similar results.

[43]In war years, over 75 percent of French revenue was spent on the military in the seventeenth century (Félix and Tallett, 2009, 155); in eighteenth century Britain, this figure varied between 61 and 74 per cent (Brewer, 1988, 32); while the peacetime military budget of Prussia during the eighteenth century accounted for 80 per cent of central government expenditure (Wilson, 2009, 119).

Figure 18: Population Estimates in China and Europe (McEvedy and Jones, 1978)



(a) Estimated population levels (b) Percentage Change

is no visible difference in population variation at the two ends of Eurasia when they were ruled by empires (before AD 400) and when they were fragmented (400–600), it is only after the consolidation of political centralization in China and fragmentation in Europe that significant differences in population change patterns emerged.

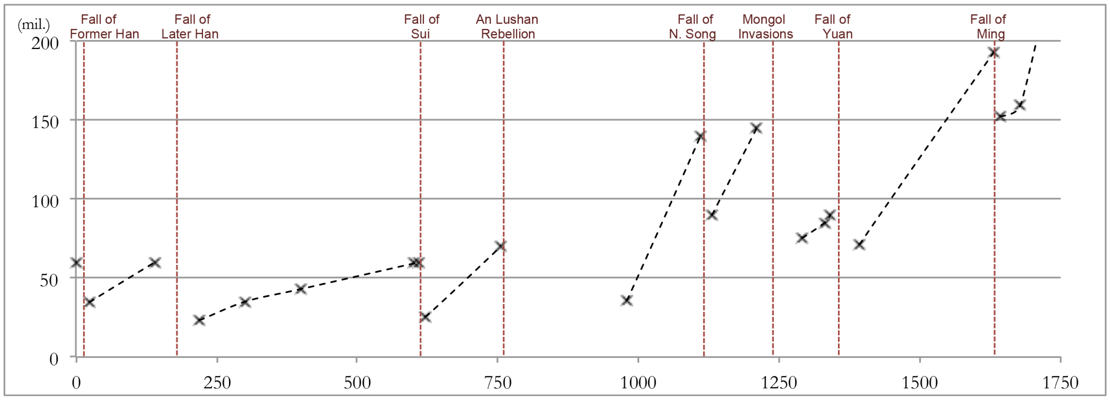
In addition, in Appendix A.7 we fit the population estimates with polynomials up to the sixth order and find that (i) it is easier to fit the European population estimates than it is to fit the Chinese population estimates because the latter are more scattered, and (ii) even if we set aside differences in the degree of goodness of fit, Europe's fitted trend line is smoother than the Chinese one.

We use McEvedy and Jones (1978) because they provide estimates for both China and Europe over a long period of time. However, since they report data for every 50, 100, or 200 years, the resulting time series is necessarily smoother than would be the case if data was available at a higher frequency. In fact, this potential problem biases us against finding a difference between the population fluctuations in China and Europe as there are several well-known sharp declines in Chinese population that are either absent or moderated in the McEvedy and Jones (1978) data.

Figure 19 displays a higher frequency population series from Cao (2000).[44] This data series is consistent with historical accounts that associate external invasions and political collapses with large declines in population. The fall of the Xin, Later Han, Sui dynasties, the An Lushan Rebellion, the fall of the Northern Song dynasty, the Mongol invasions, and the collapse of the Yuan and Ming

[44]We use the population estimates provided in Cao (2000) because of its coverage and relative accuracy. The plunges in China's population depicted in Figure 19 would be even more severe if we had used official historical statistics. For example, official historical records suggest that China's population fell from more than 50 million to 7 million in the third century after the collapse of the Han dynasty. A substantial amount of this population 'loss' was likely due to the state's inability to keep accurate records during times of crises instead of actual deaths. By contrast, Cao (2000) puts the late third-century estimate at 23 million instead of 7 million.

Figure 19: Estimated population levels and major political crises in China (Cao, 2000)



dynasties are all visible in the figure.

For example, the Mongol invasions are associated with a sharp population collapse. Kuhn observes that '[p]opulation figures took another dramatic turn downward between 1223 and 1264, and by 1292 in the whole of China the population had decreased by roughly 30 million, or one third of the population, to 75 million. This was probably due to a combination of factors—warfare in north China, the Mongol invasions, and the bubonic plague or other epidemics . . . this was a decline in human population on a magnitude that the world has seldom seen' (Kuhn, 2009, 75). The fall of the Yuan Dynasty is thought to have caused the population to fall again by approximately 23 percent. In contrast, there was only one major Europe-wide collapse in population after the fall of the Roman Empire: the Black Death of the mid-fourteenth century.

6 Conclusion

The idea that Europe's political and economic success is related to its political fragmentation goes back to the Enlightenment. Montesquieu noted that in contrast to Asia where strong nations are able to subdue their neighbors, in 'Europe on the contrary, strong nations are opposed to the strong; and those who join each other have nearly the same courage. This is the reason of the weakness of Asia and of the strength of Europe; of the liberty of Europe, and of the slavery of Asia' (Montesquieu, 1989, 266).

In this paper we have proposed a unified theory of the origins, persistence, and consequences of political centralization and fragmentation in China and Europe. We build on the argument that external threats were a powerful force for political unification in China, but were less of a factor in Europe. Our theory suggests that political centralization should indeed be stable in China, but not in Europe, and that this centralization was beneficial from a static perspective as it minimized costly interstate competition. However, we also show that in the event of an external invasion a centralized empire such as China was less robust than a decentralized state system.

Although beyond the scope of this paper, the start-stop nature of population growth that is predicted in our model has the potential to be built upon to help reconcile a big puzzle in the history of economic growth: why China, the most populous economy in the world for much of recorded history, was capable of coming 'within a hair's breadth of industrializing in the fourteenth century' (Jones, 2003, 160), but swiftly and permanently lost its technological lead after the prolonged and devastating wars of the Mongol conquests. Since more people means more ideas, growth theory often contains a scale-effect that implies that larger economies should be the first to experience modern economic growth (Kremer, 1993). However, as Aiyar et al. (2008) point out, in a (premodern) world where technological knowledge is embodied primarily in humans (instead of, for example, stored in computers), the effect of population change on the stock of knowledge is asymmetric: technological knowledge grows slowly with population growth, but regresses swiftly when the population contracts. Pairing this insight with our theory suggests that because China was more centralized and more vulnerable to negative population shocks, it experienced more frequent interruptions in cumulative innovation. In other words, China's higher variance of population growth could have diminished its chances of escaping the Malthusian trap, while the European population and economy were able to expand gradually to the point where the transition from stagnation to growth was triggered (as in theories of unified growth. See Galor and Weil 2000; Galor 2011).

In sum, our theory provides a novel channel through which geography could have shaped economic outcomes in Eurasia. Scholars have argued that decentralization gave Europe an edge in the Great Divergence because it led to greater innovation (Mokyr, 1990; Diamond, 1997; Lagerlof, 2014); support for merchants (Rosenberg and Birdzell, 1986) or political freedoms and representation (Hall, 1985). Recent work has also shown how the consequences of political fragmentation interacted with the Black Death to raise incomes and urbanization in Europe (Voigtländer and Voth, 2013b). Our theory complements these important arguments by highlighting a novel factor that contributed to political centralization in China and political fragmentation in Europe. Finally, we emphasize the significance of a previously neglected consequence of political centralization in China. There were periods of economic expansion in China, but these were brought to a halt by external invasions and political crises. It was these population crises, we conjecture, that help to explain why China did not enter a period of sustained economic growth in the preindustrial era. In contrast, Europe's polycentric system of states gave it the institutional robustness that was one of the preconditions for modern economic growth to occur.

References

Abramson, S. (2013). The economic origins of the territorial state. mimeo.

Acemoglu, D. and J. A. Robinson (2006). *The Economic Origins of Dictatorship and Democracy.* Cambridge: Cambridge University Press.

Aiyar, S., C.-J. Dalgaard, and O. Moav (2008). Technological progress and regress in pre-industrial times. *Journal of Economic Growth 13*(2), 125–144.

Alesina, A. and E. Spolaore (1997). On the number and size of nations. *Quarterly Journal of Economics 112*(4), 1027–56.

Alesina, A. and E. Spolaore (2003). *The Size of Nations.* Cambridge, MA: MIT Press.

Alesina, A. and E. Spolaore (2005). War, peace, and the size of countries. *Journal of Public Economics 89*(7), 1333–1354.

Allen, R. C., J.-P. Bassino, D. Ma, C. Moll-Murata, and J. L. V. Zanden (2011). Wages, prices, and living standards in China, 1738–1925: in comparison with Europe, Japan, and India. *Economic History Review 64*(s1), 8–38.

Baechler, J. (1975). *The Origins of Capitalism.* Oxford: Basil Blackwell.

Bai, Y. and J. Kung (2011). Climate shocks and Sino-nomadic conflict. *Review of Economics and Statistics 93*(3), 970–981.

Barfield, T. (1989). *The Perilous Frontier: Nomadic Empires and China.* Oxford: Basil Blackwell.

Bonney, R. (Ed.) (1999). *The Rise of the Fiscal State in Europe c.1200-1815.* Oxford: Oxford University Press.

Brecke, P. (1999). Violent conflicts 1400 A.D. to the present in different regions of the world.

Brewer, J. (1988). *The Sinews of Power.* Cambridge, MA: Harvard University Press.

Campante, F. R. and Q.-A. Do (2014). Isolated capital cities, accountability, and corruption: Evidence from US states. *American Economic Review 104*(8), 2456–81.

Campante, F. R., Q.-A. Do, and B. Guimaraes (2014). Capital cities, conflict, and misgovernance: Theory and evidence. Working Paper.

Cao, S. (2000). *Zhongguo ren kou shi [The Demographic history of China].* Shanghai: Fudan University Press.

Carneiro, R. L. (1970). A theory of the origin of the state. *Science 169*(3947), 733–738.

Chaliand, G. (2005). *Nomadic Empires.* New Brunswick: Transaction Books. Translated by A.M. Berrett.

Childe, V. G. (1936). *Man Makes Himself.* New York: Spokesman.

China's Military History Editorial Committee (2003). *A Chronology of Warfare in Dynastic China [zhongguo lidai zhanzheng nianbiao].* Beijing: PLA Press.

Collins, R. (1998). *Charlemagne.* Toronto: Toronto University Press.

Cowen, T. (1990). Economic effects of a conflict-prone world order. *Public Choice 64*, 121–134.

Deng, K. (2012). The continuation and efficiency of the Chinese fiscal state, 700 BC – AD 1911. In B. Yun-Casalilla, P. O'Brien, and F. Comin Comin (Eds.), *The Rise of Fiscal States: A Global History, 1500–1914.* Cambridge: Cambridge University Press.

Di Cosmo, N. (2002a). *Ancient China and its Enemies.* Cambridge: Cambridge University Press.

Di Cosmo, N. (2002b). Introduction: Inner Asian Ways of Warfare in Historical Perspective. In N. Di Cosmo (Ed.), *Warfare in Inner Asian History (500–1800)*, Leiden, pp. 1–33. Brill.

Diamond, J. (1997). *Guns, Germs, and Steel.* New York: W.W. Norton & Company.

Dincecco, M. (2009). Fiscal centralization, limited government, and public revenues in Europe, 1650–1913. *Journal of Economic History 69*(1), 48–103.

Dincecco, M. and G. Katz (2014). State capacity and long-run economic performance. *Economic Journal*. Forthcoming.

Eckstein, A. M. (2011). *Mediterranean Anarchy, Interstate War, and the Rise of Rome.* Berkeley, CA: University of California Press.

Elvin, M. (1973). *Patterns of the Chinese Past.* Stanford, CA: Stanford University Press.

Fairbank, J. K. (1992). *China: A New History.* Cambridge, MA: Belknap Press of Harvard University Press.

Félix, J. and F. Tallett (2009). The French experience, 1661-1815. In C. Storrs (Ed.), *Fiscal Military State in Eighteenth-Century Europe*, Abingdon, Oxfordshire, pp. 147–166. Ashgate.

Finer, S. (1999). *The History of Government from the Earliest Times*, Volume II. Oxford: Oxford University Press.

Food and Agriculture Organization (1984). *Livestock breeds of China (Animal Production and Health Papers vol. 46).* By Piliu Zheng. Downloadable at http://www.fao.org/docrep/012/x6549e/x6549e00.pdf.

Food and Agriculture Organization (1986). *Irrigation Water Management (FAO Training manual no. 3).* By C. Brouwer and M. Heibloem. Downloadable at ftp://ftp.fao.org/agl/aglw/fwm/Manual3.pdf.

Frachetti, M. (2008). *Pastoralist Landscapes and Social Interaction in Bronze Age Eurasia.* Berkeley: University of California Press.

Friedman, D. (1977). A theory of the size and shape of nations. *Journal of Political Economy 85*(1), 59–77.

Fukuyama, F. (2011). *The Origins of Political Order: From Prehuman Times to the French Revolution.* New York: Farrar, Straus and Giroux.

Galor, O. (2011). *Unified Growth Theory.* Princeton: Princeton University Press.

Galor, O. and D. N. Weil (2000). Population, technology, and growth: From Malthusian stagnation to the demographic transition and beyond. *American Economic Review 90*, 806–828.

Gat, A. (2006). *War in Human Civilization.* Oxford: Oxford University Press.

Gennaioli, N. and H.-J. Voth (2013). State capacity and military conflict. Technical report, memo.

Goldsworthy, A. (2009). *How Rome Fell.* New Haven: Yale University Press.

Graff, D. (2002). *Medieval Chinese Warfare, 300–900.* New York: Routledge.

Greif, A., M. F. Iyigun, and D. L. Sasson (2012). Social institutions and economic growth: Why England and not China became the first modern economy. Working Paper.

Grousset, R. (1970). *The Empire of the Steppes: a history of Central Asia.* New Brunswick: Rutgers University Press. Translated by Naomi Walford.

Gu, J. (1995). *zhongguo lidai jinian shouce [Handbook of the Annals of China's Dynasties].* Shanghai: Shanghai Ancient Books Press.

Hajnal, J. (1965). European marriage patterns in perspective. In D. Glass and D. Eversley (Eds.), *Population in History. Essays in Historical Demography*, pp. 101–143. London: Edward Arnold.

Hall, J. A. (1985). *Power and Liberties.* London: Penguin Books.

Harris, R. and R. Sollis (2003). *Applied Time Series Modelling and Forecasting.* West Sussex: Wiley.

Heather, P. (2006). *The Fall of the Roman Empire: A New History of Rome and the Barbarians.* Oxford: Oxford University Press.

Herrmann, A. (1966). *An Historical Atlas of China.* Chicago: Aldine.

Hicks, J. (1969). *A Theory of Economic History.* Oxford: Oxford University Press.

Hoffman, P. and K. Norberg (Eds.) (1994). *Fiscal Crises, Liberty, and Representative Government, 1450–1789*, Stanford, CA. Stanford University Press.

Hoffman, P. T. (2012). Why was it Europeans who conquered the world? *Journal of Economic History 72*(03), 601–633.

Huang, R. (1988). *China: A Macro History.* Armonk, NY: M.E. Sharpe.

Huang, Z. (1663, 1993). *A Plan for the Prince.* New York: Columbia University Press. Translated by Theodore de Bary.

Iyigun, M. (2008). Luther and Suleyman. *Quarterly Journal of Economics 123*(4), 1465–1494.

Johnson, N. D. and M. Koyama (2013). Legal centralization and the birth of the secular state. *Journal of Comparative Economics 41*(4), 959–978.

Johnson, N. D. and M. Koyama (2014a). Tax farming and the origins of state capacity in England and France. *Explorations in Economic History 51*(1), 1–20.

Johnson, N. D. and M. Koyama (2014b). Taxes, lawyers, and the decline of witch trials in France. *Journal of Law and Economics 57.*

Jones, E. L. (1981, 2003). *The European Miracle* (3rd ed.). Cambridge: Cambridge University Press.

Karaman, K. and Ş. Pamuk (2013). Different paths to the modern state in Europe: The interaction between warfare, economic structure and political regime. *American Political Science Review 107*(3), 603–626.

Kelly, M., J. Mokyr, and C. Ó. Gráda (2013). Precocious Albion: A new interpretation of the British Industrial Revolution. Working Papers 201311, University College Dublin.

Kiser, E. and X. Tong (1992). Determinants of the amount and type of corruption in state fiscal bureaucracies: An analysis of late imperial China. *Comparative Political Studies 25*(3), 300–331.

Ko, C. Y. and T.-H. Sng (2013). Regional dependence and political centralization in imperial China. *Eurasian Geography and Economics 54*(5–6), 470–483.

Kremer, M. (1993). Population growth and technological change: One million B.C. to 1990. *Quarterly Journal of Economics 108*(3), 681–716.

Kuhn, D. (2009). *The Age of Confucian Rule.* Cambridge, MA: Harvard University Press.

Lagerlof, N.-P. (2014). Population, technology and fragmentation: the European miracle revisited. *Journal of Development Economics 108*, 87–105.

Lattimore, O. (1940). *Inner Asian frontiers of China.* New York: American Geographical Society.

Lattimore, O. (1995 [1929]). *The desert road to Turkestan.* New York: Kodansha International.

Lewis, A. R. (1991). *Nomads and Crusaders A.D. 1000-1368.* Indianapolis: Indiana University Press.

Lieberman, V. (2009). *Strange Parallels: Southeast Asia in Global Context, C. 800-1830. Vol. 2, Mainland Mirrors: Europe, Japan, China, South Asia, and the Islands.* Cambridge: Cambridge University Press.

Ma, D. (2012). Political institutions and long run economic trajectory: Some lessons from two millennia of Chinese civilization. Technical Report 8791, CEPR.

Ma, D. (2013). State capacity and the great divergence: the case of Qing China. *Eurasian Geography and Economics 54*(56), 484–498.

McCloskey, D. N. (2010). *Bourgeois Dignity: Why Economics Can't Explain the Modern World.* Chicago: University of Chicago Press.

McEvedy, C. and R. Jones (1978). *Atlas of World Population History.* Harmondsworth: Penguin.

McKitterick, R. (2008). *Charlemagne: The formation of a European Identity.* Cambridge: Cambridge University Press.

Mencius (2004). *Mencius.* London: Penguin Books. Translated by D. C. Lau.

Mokyr, J. (1990). *The Lever of Riches.* Oxford: Oxford University Press.

Mokyr, J. (2007). The market for ideas and the origins of economic growth in eighteenth century Europe. *Tijdshrift voor Sociale en Economische Geschidenis 4*(1), 3–38.

Montesquieu, C. (1748, 1989). *The Spirit of the Laws.* Cambridge: Cambridge University Press. Translated by Anne M. Cohler, Basia C. Miller, and Harold S. Stone.

Morris, I. and W. Scheidel (Eds.) (2009). *The Dynamics of Ancient Empires: State Power from Assyria to Byzantium.* New York: Oxford University Press.

Morrissey, R. (1997). *Charlemagne & France.* Notre Dame: University of Notre Dame Press. Translated by Catherine Tihanyi.

Needham, J. (1995). *Science and Civilisation in China*, Volume VII:2 by K.G. Robinson. Cambridge: Cambridge University Press.

Ng, S. and P. Perron (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association 90*, 268–281.

Nussli, C. (2011). URL: http://www.euratlas.com/about.html.

Odahl, C. (2004). *Constantine and the Christian Empire* (2nd ed.). London: Routledge.

Olsson, O. and G. Hansson (2011). Country size and the rule of law: Resuscitating Montesquieu. *European Economic Review 55*(5), 613–629.

Perdue, P. (2005). *China Marches West: The Qing Conquest of Central Eurasia.* Cambridge: Belknap.

Pirenne, H. (1925). *Medieval Cities.* New York: Doubleday Anchor Books.

Rees, R. (2004). *Diocletian and the Tetrachy.* Edinburgh: Edinburgh University Press.

Rosenberg, N. and L. Birdzell (1986). *How the West Grew Rich: The Economic Transformation of the Industrial World.* New York: Basic Books.

Rosenthal, J.-L. and R. B. Wong (2011). *Before and Beyond Divergence.* Cambridge: Harvard University Press.

Rubin, J. (2011). Institutions, the rise of commerce and the persistence of laws: Interest restrictions in Islam and Christianity. *Economic Journal 121*(557), 1310–1339.

Scheidel, W. (2009). From the 'Great Convergence' to the 'First Great Divergence': Roman and Qin-Han state formation and its aftermath. In W. Scheidel (Ed.), *Rome and China: Comparative Perspectives on Ancient World Empires*, Oxford, pp. 11–23. Oxford University Press.

Scheidel, W. (2013). Demography. In W. Scheidel, I. Morris, and R. P. Saller (Eds.), *The Cambridge Economic history of the Greco-Roman World*, Cambridge, pp. 38–86. Cambridge University Press.

Shen, K. (2011). *Mengxi bi tan [Brush talks from dream brook].* London: Paths International. Translated by Wang Hong and Zhao Zheng.

Skinner, G. S., M. Henderson, and Z. Yue (2007). *Physiographic Macroregions.* Davis: University of California, Regional Systems Analysis Project, Jan 2007.

Sng, T.-H. (2014). Size and dynastic decline: The principal-agent problem in late imperial China 1700-1850. *Explorations in Economic History 54*(0), 107–127.

Sng, T.-H. and C. Moriguchi (2014). Asia's little divergence: State capacity in China and Japan before 1850. *Journal of Economic Growth 19*(4), 439–470.

Spruyt, H. (1994). *The Sovereign State and Its Competitors.* Princeton: Princeton University Press.

Talbert, R. (2000). *Barrington Atlas of the Greek and Roman World.* Princeton: Princeton University Press.

Tilly, C. (1990). *Coercion, Capital, and European States, AD 990-1990.* Oxford: Blackwell.

Tullock, G. (1971). The paradox of revolution. *Public Choice 11*, 89–99.

Turchin, P. (2009). A theory for formation of large empires. *Journal of Global History 4*, 191–217.

Twitchett, D. (Ed.) (1979). *The Cambridge History of China. Vol. 3: Sui and T'ang China, 589-906. Part 1.* Cambridge: Cambridge University Press.

van Creveld, M. (1977, 2004). *Supplying War: Logistics from Wallenstein to Patton.* Cambridge: Cambridge University Press.

Voigtländer, N. and H.-J. Voth (2013). Gifts of Mars: Warfare and Europe's early rise to riches. *Journal of Economic Perspectives 27*(4), 165–186.

Voigtländer, N. and H.-J. Voth (2013a). How the West invented fertility restriction. *American Economic Review 103*(6), 2227–2264.

Voigtländer, N. and H.-J. Voth (2013b). The three horsemen of riches: Plague, war, and urbanization in early modern Europe. *Review of Economic Studies 80*, 774–811.

Vries, P. (2012). Public finance in China and Britain in the long eighteenth century. Working Paper 167/12, LSE.

Ward-Perkins, B. (2005). *The Fall of Rome, and the end of civilization.* Oxford: Oxford University Press.

Watson, B. (1971). *Records of the grand historian of China Vol. 2.* New York: Columbia University Press.

Wei, L. (2011). *zhongguo wenshi jianbiao huibian [Concise tables of Chinese cultural and historical materials].* Beijing: Commercial Press.

White, M. (2013). *Atrocities: The 100 Deadliest Episodes in Human History.* New York: W. W. Norton & Company.

Wilkinson, E. (2012). *Chinese history: A new manual.* Cambridge, MA: Harvard University Asia Center.

Wilson, P. (2009). Prussia as a fiscal-military state, 1640-1806. In C. Storrs (Ed.), *Fiscal Military State in Eighteenth-Century Europe*, Abingdon, Oxfordshire, pp. 95–124. Ashgate.

Zhang, B., F. Feng, L. Zhang, and H. Li (1994). *zhongguo nongye ziran zaihai shiliao ji [A Compendium of Historical Materials on the Natural Disasters in Chinese Agriculture].* Xi'an: Shaanxi Science & Technology Press.

Zhang, D., Q. Pei, H. Lee, J. Zhang, C. Q. Chang, B. Li, J. Li, and X. Zhang (2015). The pulse of imperial China: a quantitative analysis of long-term geopolitical and climatic cycles. *Global Ecology and Biogeography 24*(1), 87–96.

(For Online Publication)

A Appendix

A.1 Eurasia's External Threats

Phase	Century	Nomadic Peoples	W. Europe	Russia	China
1st	8th–2nd BC	Scythians		✓	
2nd	2nd BC	Xiongnu (Hsiung-nu)			✓
	2nd BC	Yuezhi (Yüeh-chih)			✓
3rd	4th	To-pa (Toba Turks)			✓
	5th	Huns	✓	✓	
	6th	Tu-chueh (Göktũrks)			✓
	6th	Ruren (Juanjuan)			✓
4th	7th	Avars	✓	✓	
	8th	Bulgars	✓		
	7th	Khazar Turks		✓	
	9th	Magyars	✓		
	9th	Uygurs			✓
	10th	Khitans			✓
5th	11th	Pechenegs and Kipchaks		✓	
	12th	Jurchens			✓
6th	13th	Mongols		✓	✓
	14th	Tatars		✓	✓
7th	15th	Oirots			✓
	17th	Manchus			✓

Table 5: Major waves of nomadic invasions. Source: Chaliand (2005). It is evident that China faced a greater threat from the steppe invaders than did Europe. See Section 2 for a historical discussion.

Invader	Date of Invasion	Location of Invasion	Direction of threat
Huns	c. 370–450	Italy, France, Balkans	East
Avars	580	South-Eastern Europe	East
Bulgars	c. 850	South-Eastern Europe	East
Arabs	711	Spain	South
Arabs	721	France	South
Arabs	732	France	South
Vikings*	793–1066	Britain	North
Vikings*	c. 810–1000	France	North
Vikings*	c. 810–1000	Low Countries	North
Arabs	831	Sicily	South
Arabs	840	Crete	South
Arabs	846	Italy	South
Magyars	907	Germany	East
Magyars	917	France	East
The Almohads	1172	Spain	South
Mongols	1240	Poland	East
Mongols	1241	Hungary	East
Mongols	1241	Croatia	East
The Marinids	1340	Gibraltar	South
Ottomans	1371	Serbia	South-East
Ottomans	1385	Albania	South-East
Ottomans	1463	Bosnia	South-East
Ottomans	1479	Hungary	South East
Ottomans	1480	Italy	South East
Crimean Tatars	1480–1507	Poland	East
Crimean Tatars	1507–1570	Russia	East
Ottomans	1526	Hungary	South-East
Ottomans	1529	Austria	South-East
Ottomans	1541	Hungary	South-East
Ottomans	1565	Malta	South-East
Ottomans	1573	Cyprus	South-East
Crimean Tatars	1571	Russia	East
Crimean Tatars	1577, 1584, 1590	Poland	East
Ottomans	1573	Cyprus	South-East
Crimean Tatars	1621	Poland	East
Crimean Tatars	1648	Poland	East
Crimean Tatars	1667	Poland	East
Ottomans	1669	Crete	South-East
Ottomans	1683	Austria	South-East

Table 6: The Multidirectional Threat in Europe. A list of invasions of Europe from the Fall of the Roman Empire onward. See Section 2 for a historical discussion. We list invasion attempts that failed as well as those that succeeded. *We count the Vikings as external invaders—due to their different religion and distinct culture they were seen as outsider invaders by contemporaries. But from the perspective of our model, we consider later Swedish or Danish campaigns in Europe as instances of interstate competition.

A.2 Proofs of Propositions in Section 3

Here, we provide the proofs of Propositions 1–3.

A.2.1, A.2.2, and A.2.3 correspond respectively to Propositions 1, 2, and 3. In A.2.4, we add plausible assumptions on the shape of the border and military investment functions to reinterpret Proposition 3.

A.2.1 Proof of Proposition 1

Proof. First, consider a two-sided threat.

1. If $\Lambda \leq \Lambda_I$, the revolution constraint is not violated even if military investment is zero. Since regime e's payoff is decreasing in its military investment, it is optimal to invest zero regardless of the location of the capital city.

2–3. Given that the two-sided threat is symmetric, there is no loss of generality if we assume $G_e \leq 1/2$. Since regime e always receives gross tax revenue y, maximizing its net tax revenue is equivalent to minimizing its military expenditure:

$$\min_{G_e, M_e} c\left(M_e\right)$$

subject to

$$\begin{aligned} \lambda\left(\Lambda, a\right) &\leq m\left(M_e, G_e - a, \beta\right), \\ \lambda\left(\Lambda, 1-z\right) &\leq m\left(M_e, z - G_e, \beta\right), \\ z - a &\geq \delta, \text{and} \\ a \leq G_e &\leq z. \end{aligned}$$

where a and z are, respectively, the leftmost and rightmost locations in the empire that suffers zero net damage when the military investment is M_e. If $M_e = 0$ satisfies every inequality, then the optimal solution is a corner one where regime e invests zero in the military. This is the case when $\Lambda \leq \Lambda_I$ (as discussed above). Now consider the case when the solution is interior with a positive military investment, i.e. $M_e > 0$. Note that the first inequality must bind in equilibrium. Otherwise, regime e can increase its net tax revenue by reducing M_e and increasing G_e. Hence, we have:

$$\lambda\left(\Lambda, a\right) = m\left(M_e, G_e - a, \beta\right).$$

Next consider the two cases when (i) the second inequality is binding, and (ii) when it does not bind. Case (i): $\lambda(\Lambda, 1-z) = m(M_e, z - G_e, \beta)$ because Λ exceeds some threshold (since $\partial\lambda/\partial\Lambda > 0$). Let Λ_{II} denote this threshold (which we will define it later). In this case ($\Lambda \geq \Lambda_{II}$), the third inequality automatically binds, i.e., $z = a + \delta$. Hence, we have:

$$\lambda(\Lambda, a) = m(M_e^*, G_e^* - a, \beta), \text{and}$$
$$\lambda(\Lambda, 1-\delta-a) = m(M_e^*, \delta + a - G_e^*, \beta).$$

Applying total differentiation:

$$\begin{bmatrix} \lambda_2(\Lambda, a) + m_2(M_e^*, G_e^* - a, \beta) & -m_1(M_e^*, G_e^* - a, \beta) \\ -\lambda_2(\Lambda, 1-\delta-a) - m_2(M_e^*, \delta + a - G_e^*, \beta) & -m_1(M_e^*, \delta + a - G_e^*, \beta) \end{bmatrix} \begin{bmatrix} da \\ dM_e \end{bmatrix}$$
$$= \begin{bmatrix} -\lambda_1(\Lambda, a) \\ -\lambda_1(\Lambda, 1-\delta-a) \end{bmatrix} d\Lambda + \begin{bmatrix} m_2(M_e^*, G_e^* - a, \beta) \\ -m_2(M_e^*, \delta + a - G_e^*, \beta) \end{bmatrix} dG_e.$$

It is easy to show that

$$\Delta = \begin{vmatrix} \lambda_2(\Lambda, a) + m_2(M_e^*, G_e^* - a, \beta) & -m_1(M_e^*, G_e^* - a, \beta) \\ -\lambda_2(\Lambda, 1-\delta-a) - m_2(M_e^*, \delta + a - G_e^*, \beta) & -m_1(M_e^*, \delta + a - G_e^*, \beta) \end{vmatrix}$$
$$> 0 \text{ because } m_1 > 0,\ \lambda_2 < 0 \text{ and } m_2 < 0.$$

Hence

$$\frac{dM_e}{dG_e} = \frac{\begin{vmatrix} \lambda_2(\Lambda, a) + m_2(M_e^*, G_e^* - a, \beta) & -m_2(M_e^*, G_e^* - a, \beta) \\ -\lambda_2(\Lambda, 1-\delta-a) - m_2(M_e^*, \delta + a - G_e^*, \beta) & -m_2(M_e^*, \delta + a - G_e^*, \beta) \end{vmatrix}}{\Delta}$$
$$< 0 \text{ because } \lambda_2 < 0 \text{ and } m_2 < 0.$$

Since $\partial c/\partial M_e > 0$ and therefore $\partial V_e/\partial M_e < 0$, we have

$$\frac{dV_e}{dG_e} = \frac{\partial V_e}{\partial M_e}\frac{dM_e}{dG_e} > 0.$$

Hence, to maximize its net tax revenue, regime e should locate its capital as close to $1/2$ as possible. This implies that $G_e^* = 1/2$, and $a = \frac{1}{2}(1-\delta)$. The optimal military spending M_e^* satisfies:

$$\lambda\left(\Lambda, \frac{1}{2}(1-\delta)\right) = m(M_e^*, \delta/2, \beta).$$

Note that case (i) denotes case 3 of Proposition 1.

Case (ii): if $\lambda(\Lambda, 1-z) < m(M_e, z-G_e)$, then $x^*(\Lambda) < 1/2$, which in turn implies that:

$$z = 1 - x^*(\Lambda).$$

Otherwise, regime e can increase its net tax revenue by reducing M_e and G_e simultaneously. By the same reasoning, the third inequality must bind in equilibrium:

$$G_e^* = a = z - \delta.$$

Therefore, the optimal military spending M_e^* must satisfy:

$$\lambda(\Lambda, 1 - x^*(\Lambda) - \delta) = m(M_e^*, 0, \beta).$$

Case (ii) denotes case 2 of Proposition 1. The threshold Λ_{II} is the solution to the following system of equations:

$$\lambda\left(\Lambda_{II}, \frac{1}{2}(1-\delta)\right) = m(M_e^*, \delta/2, \beta), \text{ and}$$
$$\lambda(\Lambda_{II}, 1 - x^*(\Lambda_{II}) - \delta) = m(M_e^*, 0, \beta).$$

Finally, consider a one-sided threat.

4. If $\Lambda \leq \bar{\Lambda}_I$, then $x^*(\Lambda) \leq 1-\delta$ so that the fraction of protected area is no less than δ even if there is no military investment. Since regime e's payoff is decreasing in its military investment, the optimal military investment is zero and the capital city is located between 0 and 1.

5. If $x^*(\Lambda) > 1-\delta$, regime e has to make a strictly positive military investment. Since military strength decreases over distance ($m_2 < 0$), it should locate its capital city at the point where the revolution constraint just binds. This implies that $G_e^* = 1-\delta$ and M_e^* solves $\lambda(\Lambda, 1-\delta) = m(M_e^*, 0, \beta)$. □

A.2.2 Proof of Proposition 2

Before proving the proposition, it is useful to characterize the outcome of interstate competition in the absence of external threats ($\Lambda = 0$).

Lemma 1. *When there is no external threat, given some locations of capital cities G_l and G_r, the* *equilibrium military investments M_l^* and M_r^* satisfy:*

$$\frac{\partial b}{\partial M_l}y - \frac{\partial c}{\partial M_l} = 0, \textit{ and}$$
$$-\frac{\partial b}{\partial M_r}y - \frac{\partial c}{\partial M_r} = 0.$$

For any symmetric equilibrium capital locations G_l^* and G_r^* that satisfy $G_l^* \neq 0$ and $G_r^* \neq 1$ (i.e., interior solutions), we have

$$\left(\frac{\partial b}{\partial G_l} + \frac{\partial b}{\partial M_l}\frac{\partial M_l}{\partial G_l} + \frac{\partial b}{\partial M_r}\frac{\partial M_r}{\partial G_l}\right)y - \frac{\partial c}{\partial M_l}\frac{\partial M_l}{\partial G_l} = 0, \textit{ and}$$
$$-\left(\frac{\partial b}{\partial G_r} + \frac{\partial b}{\partial M_l}\frac{\partial M_l}{\partial G_r} + \frac{\partial b}{\partial M_r}\frac{\partial M_r}{\partial G_r}\right)y - \frac{\partial c}{\partial M_r}\frac{\partial M_r}{\partial G_r} = 0.$$

Proof. First consider the second stage of interstate competition. Given G_l and G_r, the optimization problems for regimes l and r are

$$\max_{M_l} V_l = b\left(G_l, G_r, M_l, M_r, \beta\right)y - c\left(M_l\right), \text{ and}$$
$$\max_{M_r} V_r = (1 - b\left(G_l, G_r, M_l, M_r, \beta\right))y - c\left(M_r\right).$$

The respective FOCs are

$$\frac{\partial b}{\partial M_l}y - \frac{\partial c}{\partial M_l} = 0, \text{ and}$$
$$-\frac{\partial b}{\partial M_r}y - \frac{\partial c}{\partial M_r} = 0,$$

Given the setup, it cannot be an equilibrium for any regime under interstate competition to invest zero in the military, i.e., it must be the case that $M_l^* > 0$ and $M_r^* > 0$. Since $\partial^2 c/\partial^2 M_l < 0$ and $\partial^2 c/\partial^2 M_r < 0$, the SOCs are guaranteed if $\partial^2 b/\partial M_l^2 \leq 0$ and $\partial^2 b/\partial M_r^2 \geq 0$.

The second-stage equilibrium military investments by regime l and regime r are $M_l^*\left(G_l, G_r\right)$ and $M_r^*\left(G_l, G_r\right)$ respectively. Let $b^*\left(G_l, G_r, \beta\right) \equiv b\left(G_l, G_r, M_l^*\left(G_l, G_r\right), M_r^*\left(G_l, G_r\right), \beta\right)$, $c_l^*\left(G_l, G_r\right) \equiv c\left(M_l^*\left(G_l, G_r\right)\right)$ and $c_r^*\left(G_l, G_r\right) \equiv c\left(M_r^*\left(G_l, G_r\right)\right)$.

Consider the first stage where regimes l and r decide their capital city locations:

$$\max_{G_l} V_l = b^*\left(G_l, G_r, \beta\right)y - c_l^*\left(G_l, G_r\right), \text{ and}$$
$$\max_{G_r} V_r = (1 - b^*\left(G_l, G_r, \beta\right))y - c_r^*\left(G_l, G_r\right).$$

The respective FOCs are

$$\frac{\partial b}{\partial G_l} y - \frac{\partial c_l}{\partial G_l} = 0, \text{ and}$$
$$-\frac{\partial b}{\partial G_r} y - \frac{\partial c_r}{\partial G_r} = 0$$

Hence, we have

$$\left(\frac{\partial b}{\partial G_l} + \frac{\partial b}{\partial M_l} \frac{\partial M_l}{\partial G_l} + \frac{\partial b}{\partial M_r} \frac{\partial M_r}{\partial G_l} \right) y - \frac{\partial c}{\partial M_l} \frac{\partial M_l}{\partial G_l} = 0, \text{ and}$$
$$-\left(\frac{\partial b}{\partial G_r} + \frac{\partial b}{\partial M_l} \frac{\partial M_l}{\partial G_r} + \frac{\partial b}{\partial M_r} \frac{\partial M_r}{\partial G_r} \right) y - \frac{\partial c}{\partial M_r} \frac{\partial M_r}{\partial G_r} = 0.$$

□

Proof. First, we denote the equilibrium in the absence of external threats by $M_l^* = M_r^* = M^*$ and $G_l^* = 1 - G_r^* = G^*$.

Now consider a two-sided threat. Let $\Lambda_{III}(\beta)$ solve:

$$\lambda \left(\Lambda_{III}(\beta), \frac{1}{2} (1 - \delta) \right) = m \left(M^*, G^* - \frac{1}{2} (1 - \delta), \beta \right).$$

It is clear that if $\Lambda \leq \Lambda_{III}(\beta)$, the revolution constraint of the two regimes does not bind, and vice versa.

Next, consider a one-sided threat. Let $\bar{\Lambda}_{III}(\beta)$ solve:

$$\lambda \left(\bar{\Lambda}_{III}(\beta), 1 - \delta \right) = m \left(M^*, G^* - (1 - \delta), \beta \right).$$

It is clear that if $\bar{\Lambda} \leq \Lambda_{III}(\beta)$, the revolution constraint of the two regimes does not bind, and vice versa. □

A.2.3 Proof of Proposition 3

Before proving the proposition, it is useful to characterize the outcome of interstate competition when the external threat is two-sided.

Lemma 2. *For a two-sided external threat where $\Lambda \leq \Lambda_{III}(\beta)$, the equilibrium outcome coincides* *with the case when there is no external threat. If $\Lambda > \Lambda_{III}(\beta)$, in the symmetric equilibrium,*

$$\frac{\partial b}{\partial M_l}y - \frac{\partial c}{\partial M_l} < 0,$$
$$-\frac{\partial b}{\partial M_r}y - \frac{\partial c}{\partial M_r} < 0.$$

In addition,

$$\frac{dM_l}{d\beta} = \frac{dM_r}{d\beta} \geq 0,\ \frac{\partial G_l}{\partial \beta} = -\frac{\partial G_r}{\partial \beta} \leq 0,$$

where equality holds when $G_l^* = \frac{1}{2}(1-\delta)$ and $G_r^* = \frac{1}{2}(1+\delta)$.

Proof. When $\Lambda \leq \Lambda_{III}(\beta)$, the revolution constraint is not binding so that the equilibrium in the absence of external threat remains an equilibrium. When $\Lambda > \Lambda_{III}(\beta)$, the revolution constraint binds, which implies that M_l and M_r are greater than the solutions when the threat is absent for any given G_l and G_r. Since $\frac{\partial^2 b}{\partial M_l^2} \leq 0$ and $\frac{\partial^2 c}{\partial M_l^2} > 0$, it follows that $\frac{\partial b}{\partial M_l}y - \frac{\partial c}{\partial M_l} < 0$. Similarly, since $\frac{\partial^2 b}{\partial M_r^2} \geq 0$ and $\frac{\partial^2 c}{\partial M_r^2} > 0$, it follows that $-\frac{\partial b}{\partial M_r}y - \frac{\partial c}{\partial M_r} < 0$.

Given that the revolution constraint binds, regimes l and r together protect δ fraction of the continent. By symmetry, each regime has to protect $\delta/2$ fraction of the interval.

We first argue that $G_l^* \geq \frac{1}{2}(1-\delta)$ and $G_r^* \leq \frac{1}{2}(1-\delta)$. Suppose to the contrary, $G_l^* < \frac{1}{2}(1-\delta)$. By symmetry, $G_r^* = 1 - G_l^* > \frac{1}{2}(1+\delta)$. Let $\varepsilon > 0$, if regime l moves to $G_l^* + \varepsilon < \frac{1}{2}(1-\delta)$ so that the revolution constraint remains binding, regime l can reduce M_l and hence increase its net tax revenue, which contradicts that this is an equilibrium. Therefore, $G_l^* \geq \frac{1}{2}(1-\delta)$ and $G_r^* \leq \frac{1}{2}(1-\delta)$.

Now the constraints are $h_l \geq 0$ and $h_r \geq 0$ where $h_l \equiv m\left(M_l, G_l - \frac{1}{2}(1-\delta), \beta\right) - \lambda\left(\Lambda, \frac{1}{2}(1-\delta)\right)$ and $h_r \equiv m\left(M_r, (1-G_r) - \frac{1}{2}(1-\delta), \beta\right) - \lambda\left(\Lambda, \frac{1}{2}(1-\delta)\right)$.

Hence,

$$\frac{\partial h_l}{\partial M_l} = m_1, \frac{\partial h_r}{\partial M_l} = 0, \frac{\partial h_l}{\partial M_r} = 0,\ \text{and}\ \frac{\partial h_r}{\partial M_r} = m_1,$$

and

$$\frac{\partial h_l}{\partial G_l} = -\frac{\partial h_r}{\partial G_r} = m_2, \frac{\partial h_r}{\partial G_l} = \frac{\partial h_l}{\partial G_r} = 0,\ \text{and}\ \frac{\partial h_l}{\partial \beta} = \frac{\partial h_r}{\partial \beta} = m_3.$$

In addition,

$$\frac{\partial^2 h_l}{\partial G_l^2} = \frac{\partial^2 h_r}{\partial G_r^2} = m_{22}, \frac{\partial^2 h_r}{\partial G_l^2} = \frac{\partial^2 h_r}{\partial G_l^2} = 0, \frac{\partial^2 h_l}{\partial G_l \partial M_l} = \frac{\partial^2 h_r}{\partial G_r \partial M_r} = -m_{12},$$
$$\frac{\partial^2 h_l}{\partial G_l \partial M_r} = \frac{\partial^2 h_l}{\partial G_r \partial M_r} = \frac{\partial^2 h_l}{\partial G_r \partial M_l} = \frac{\partial^2 h_r}{\partial G_l \partial M_r} = \frac{\partial^2 h_r}{\partial G_l \partial M_l} = \frac{\partial^2 h_r}{\partial G_r \partial M_l} = 0,$$
$$\frac{\partial^2 h_r}{\partial \beta \partial G_l} = \frac{\partial^2 h_l}{\partial \beta \partial G_r} = \frac{\partial^2 h_l}{\partial \beta \partial M_r} = \frac{\partial^2 h_r}{\partial \beta \partial M_l} = 0,$$
$$\frac{\partial^2 h_l}{\partial \beta \partial G_l} = -\frac{\partial^2 h_r}{\partial \beta \partial G_r} = m_{23}, \text{ and } \frac{\partial^2 h_l}{\partial \beta \partial M_l} = \frac{\partial^2 h_r}{\partial \beta \partial M_r} = m_{13}.$$

Note that M_l and M_r are determined separately by the constraints $h_l = 0$ and $h_r = 0$. Total differentiation gives,

$$\frac{dM_l}{dG_l} = \frac{-\frac{\partial h_l}{\partial G_l}}{\frac{\partial h_l}{\partial M_l}} = -\frac{m_2}{m_1}, \frac{dM_r}{dG_r} = \frac{-\frac{\partial h_r}{\partial G_r}}{\frac{\partial h_r}{\partial M_r}} = \frac{m_2}{m_1},$$
$$\frac{dM_l}{d\beta} = \frac{-\frac{\partial h_l}{\partial \beta}}{\frac{\partial h_l}{\partial M_l}} = \frac{-m_3}{m_1}, \frac{dM_r}{d\beta} = -\frac{\frac{\partial h_r}{\partial \beta}}{\frac{\partial h_r}{\partial M_r}} = \frac{-m_3}{m_1}, \text{ and}$$
$$\frac{dM_r}{dG_l} = \frac{dM_l}{dG_r} = 0.$$

When $G_l = 1 - G_r$, it follows that $\frac{dM_l}{dG_l} = -\frac{dM_r}{dG_r}$, $\frac{dM_l}{d\beta} = \frac{dM_r}{d\beta}$, and

$$\frac{dM_l}{dG_l} + \frac{dM_r}{dG_l} = -\left(\frac{dM_r}{dG_r} + \frac{dM_l}{dG_r}\right) = -\frac{m_2}{m_1}.$$

In addition,

$$\frac{d^2 M_l}{d\beta^2} = \frac{m_1\left(-m_{33}\right) + m_3 m_{13}}{m_1^2} \geq 0, \frac{d^2 M_r}{d\beta^2} = \frac{m_1\left(-m_{33}\right) + m_3 m_{13}}{m_1^2} \geq 0.$$

Furthermore,

$$\frac{d^2M_r}{dG_l^2} = \frac{d^2M_l}{dG_r^2} = 0,$$

$$\frac{d^2M_l}{dG_l^2} = \frac{\left(\frac{\partial h_l}{\partial M_l}\right)\left(-\frac{\partial^2 h_l}{\partial G_l^2}\right) + \frac{\partial h_l}{\partial G_l}\frac{\partial^2 h_l}{\partial G_l \partial M_l}}{\left(\frac{\partial h_l}{\partial M_l}\right)^2} = \frac{-m_1 m_{22} + m_2 m_{12}}{(m_1)^2} > 0,$$

$$\frac{d^2M_r}{dG_r^2} = \frac{\frac{\partial h_r}{\partial M_r}\left(-\frac{\partial^2 h_r}{\partial G_r^2}\right) + \frac{\partial h_r}{\partial G_r}\frac{\partial^2 h_r}{\partial G_r \partial M_r}}{\left(\frac{\partial h_r}{\partial M_r}\right)^2} = \frac{-m_1 m_{22} + m_2 m_{12}}{(m_1)^2} > 0,$$

$$\frac{d^2M_l}{dG_l dG_r} = \frac{\left(\frac{\partial h_l}{\partial M_l}\right)\left(-\frac{\partial^2 h_l}{\partial G_r \partial G_l}\right) + \frac{\partial h_l}{\partial G_l}\frac{\partial^2 h_l}{\partial G_r \partial M_l}}{\left(\frac{\partial h_l}{\partial M_l}\right)^2} = 0,$$

$$\frac{d^2M_r}{dG_l dG_r} = \frac{\frac{\partial h_r}{\partial M_r}\left(-\frac{\partial^2 h_r}{\partial G_r \partial G_l}\right) + \frac{\partial h_r}{\partial G_r}\frac{\partial^2 h_r}{\partial G_l \partial M_r}}{\left(\frac{\partial h_r}{\partial M_r}\right)^2} = 0,$$

$$\frac{d^2M_l}{dG_l d\beta} = \frac{\frac{\partial h_l}{\partial M_l}\left(-\frac{\partial^2 h_l}{\partial \beta \partial G_l}\right) + \left(\frac{\partial h_l}{\partial G_l}\right)\frac{\partial^2 h_l}{\partial \beta \partial M_l}}{\left(\frac{\partial h_l}{\partial M_l}\right)^2} = \frac{-m_1 m_{23} + m_2 m_{13}}{m_1^2} > 0,$$

$$\frac{d^2M_l}{dG_r d\beta} = \frac{d^2M_r}{dG_l d\beta} = 0,$$

$$\frac{d^2M_r}{dG_r d\beta} = \frac{\frac{\partial h_r}{\partial M_r}\left(-\frac{\partial^2 h_r}{\partial \beta \partial G_r}\right) + \frac{\partial h_r}{\partial G_r}\frac{\partial^2 h_r}{\partial \beta \partial M_r}}{\left(\frac{\partial h_r}{\partial M_r}\right)^2} = \frac{m_1 m_{23} - m_2 m_{13}}{m_1^2} < 0.$$

The second-stage equilibrium military investments by regime l and regime r are $M_l^*(G_l, G_r)$ and $M_r^*(G_l, G_r)$ respectively. Let $b^*(G_l, G_r, \beta) \equiv b(G_l, G_r, M_l^*(G_l, G_r), M_r^*(G_l, G_r), \beta)$, $c_l^*(G_l, G_r) \equiv c(M_l^*(G_l, G_r))$ and $c_r^*(G_l, G_r) \equiv c(M_r^*(G_l, G_r))$. Now,

$$
\begin{aligned}
\frac{\partial^2 b^*}{\partial G_l^2} &= \frac{\partial^2 b}{\partial G_l^2} + \frac{\partial^2 b}{\partial M_l^2}\left(\frac{\partial M_l}{\partial G_l}\right)^2 + \frac{\partial^2 b}{\partial M_r^2}\left(\frac{\partial M_r}{\partial G_l}\right)^2 + \frac{\partial b}{\partial M_l}\frac{\partial^2 M_l}{\partial G_l^2} + \frac{\partial b}{\partial M_r}\frac{\partial^2 M_r}{\partial G_l^2} \\
&= \frac{\partial^2 b}{\partial G_l^2} + \frac{\partial^2 b}{\partial M_l^2}\left(-\frac{m_2}{m_1}\right)^2 + \frac{\partial b}{\partial M_l}\left(\frac{-m_1 m_{22} + m_2 m_{12}}{(m_1)^2}\right), \\
\frac{\partial^2 b^*}{\partial G_r^2} &= \frac{\partial^2 b}{\partial G_r^2} + \frac{\partial^2 b}{\partial M_l^2}\left(\frac{\partial M_l}{\partial G_r}\right)^2 + \frac{\partial^2 b}{\partial M_r^2}\left(\frac{\partial M_r}{\partial G_r}\right)^2 + \frac{\partial b}{\partial M_l}\frac{\partial^2 M_l}{\partial G_r^2} + \frac{\partial b}{\partial M_r}\frac{\partial^2 M_r}{\partial G_r^2} \\
&= \frac{\partial^2 b}{\partial G_r^2} + \frac{\partial^2 b}{\partial M_r^2}\left(-\frac{m_2}{m_1}\right)^2 + \frac{\partial b}{\partial M_r}\left(\frac{-m_1 m_{22} + m_2 m_{12}}{(m_1)^2}\right), \\
\frac{\partial^2 b^*}{\partial G_r \partial G_l} &= \frac{\partial^2 b}{\partial G_r \partial G_l} + \frac{\partial^2 b}{\partial M_l^2}\frac{\partial M_l}{\partial G_l}\frac{\partial M_l}{\partial G_r} + \frac{\partial^2 b}{\partial M_r^2}\frac{\partial M_r}{\partial G_l}\frac{\partial M_r}{\partial G_r} + \frac{\partial b}{\partial M_l}\frac{\partial^2 M_l}{\partial G_r \partial G_l} + \frac{\partial b}{\partial M_r}\frac{\partial^2 M_r}{\partial G_r \partial G_l} \\
&= \frac{\partial^2 b}{\partial G_r \partial G_l}, \\
\frac{db^*}{dG_l d\beta} &= \frac{\partial^2 b}{\partial \beta \partial G_l} + \frac{\partial^2 b}{\partial M_l^2}\frac{\partial M_l}{\partial G_l}\frac{\partial M_l}{\partial \beta} + \frac{\partial^2 b}{\partial M_r^2}\frac{\partial M_r}{\partial G_l}\frac{\partial M_r}{\partial \beta} + \frac{\partial b}{\partial M_l}\frac{\partial^2 M_l}{\partial \beta \partial G_l} + \frac{\partial b}{\partial M_r}\frac{\partial^2 M_r}{\partial \beta \partial G_l} \\
&= \frac{\partial^2 b}{\partial \beta \partial G_l} + \frac{\partial^2 b}{\partial M_l^2}\frac{\partial M_l}{\partial G_l}\frac{\partial M_l}{\partial \beta} + \frac{\partial b}{\partial M_l}\frac{\partial^2 M_l}{\partial \beta \partial G_l} \\
&= \frac{\partial^2 b}{\partial \beta \partial G_l} + \frac{\partial^2 b}{\partial M_l^2}\frac{m_2}{m_1}\frac{m_3}{m_1} + \frac{\partial b}{\partial M_l}\left(\frac{-m_1 m_{23} + m_2 m_{13}}{m_1^2}\right), \\
\frac{db^*}{dG_r d\beta} &= \frac{\partial^2 b}{\partial \beta \partial G_l} + \frac{\partial^2 b}{\partial M_l^2}\frac{\partial M_l}{\partial G_r}\frac{\partial M_l}{\partial \beta} + \frac{\partial^2 b}{\partial M_r^2}\frac{\partial M_r}{\partial G_r}\frac{\partial M_r}{\partial \beta} + \frac{\partial b}{\partial M_l}\frac{\partial^2 M_l}{\partial \beta \partial G_r} + \frac{\partial b}{\partial M_r}\frac{\partial^2 M_r}{\partial \beta \partial G_r} \\
&= \frac{\partial^2 b}{\partial \beta \partial G_r} + \frac{\partial^2 b}{\partial M_r^2}\frac{\partial M_r}{\partial G_r}\frac{\partial M_r}{\partial \beta} + \frac{\partial b}{\partial M_r}\frac{\partial^2 M_r}{\partial \beta \partial G_r} \\
&= \frac{\partial^2 b}{\partial \beta \partial G_r} + \frac{\partial^2 b}{\partial M_r^2}\left(-\frac{m_2}{m_1}\right)\frac{m_3}{m_1} + \frac{\partial b}{\partial M_r}\left(-\frac{-m_1 m_{23} + m_2 m_{13}}{m_1^2}\right),
\end{aligned}
$$

$$
\begin{aligned}
\frac{d^2c_l^*}{dG_l^2} &= \frac{d^2c}{dM_l^2}\left(\frac{\partial M_l}{\partial G_l}\right)^2 + \frac{dc}{dM_l}\frac{\partial^2 M_l}{\partial G_l^2} \\
&= \frac{d^2c}{dM_l^2}\left(-\frac{m_2}{m_1}\right)^2 + \frac{dc}{dM_l}\left(\frac{-m_1m_{22}+m_2m_{12}}{(m_1)^2}\right) > 0, \\
\frac{d^2c_r^*}{dG_r^2} &= \frac{d^2c}{dM_r^2}\left(\frac{\partial M_r}{\partial G_r}\right)^2 + \frac{dc}{dM_r}\frac{\partial^2 M_r}{\partial G_r^2} \\
&= \frac{d^2c}{dM_r^2}\left(-\frac{m_2}{m_1}\right)^2 + \frac{dc}{dM_r}\left(\frac{-m_1m_{22}+m_2m_{12}}{(m_1)^2}\right) > 0, \\
\frac{d^2c_l^*}{dG_lG_r} &= \frac{d^2c}{dM_l^2}\frac{\partial M_l}{\partial G_l}\frac{\partial M_l}{\partial G_r} + \frac{dc}{dM_l}\frac{\partial^2 M_l}{\partial G_l\partial G_r} = 0, \\
\frac{d^2c_r^*}{dG_lG_r} &= \frac{d^2c}{dM_r^2}\frac{\partial M_r}{\partial G_r}\frac{\partial M_r}{\partial G_l} + \frac{dc}{dM_r}\frac{\partial^2 M_r}{\partial G_l\partial G_r} = 0,
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{d^2c_l^*}{dG_ld\beta} &= \frac{d^2c}{dM_l^2}\frac{\partial M_l}{\partial G_l}\frac{\partial M_l}{\partial \beta} + \frac{dc}{dM_l}\frac{\partial^2 M_l}{\partial G_l\partial \beta} \\
&= \frac{d^2c}{dM_l^2}\left(-\frac{m_2}{m_1}\right)\left(\frac{-m_3}{m_1}\right) + \frac{dc}{dM_l}\left(\frac{-m_1m_{23}+m_{23}m_{13}}{m_1^2}\right) \geq 0, \\
\frac{d^2c_r^*}{dG_rd\beta} &= \frac{d^2c}{dM_r^2}\frac{\partial M_r}{\partial G_r}\frac{\partial M_r}{\partial \beta} + \frac{dc}{dM_r}\frac{\partial^2 M_r}{\partial G_r\partial \beta} \\
&= \frac{d^2c}{dM_r^2}\left(\frac{m_2}{m_1}\right)\left(\frac{-m_3}{m_1}\right) + \frac{dc}{dM_r}\frac{m_1m_{23}-m_2m_{13}}{m_1^2} \leq 0.
\end{aligned}
$$

Now,

$$
\begin{aligned}
& -\frac{db^*}{dG_ld\beta}y + \frac{d^2c_l^*}{dG_ld\beta} \\
&= -\frac{\partial^2 b}{\partial\beta\partial G_l}y - \frac{\partial^2 b}{\partial M_l^2}\frac{m_2}{m_1}\frac{m_3}{m_1}y - \frac{\partial b}{\partial M_l}\left(\frac{-m_1m_{23}+m_2m_{13}}{m_1^2}\right)y \\
&+ \frac{d^2c}{dM_l^2}\left(-\frac{m_2}{m_1}\right)\left(\frac{-m_3}{m_1}\right) + \frac{dc}{dM_l}\left(\frac{-m_1m_{23}+m_{23}m_{13}}{m_1^2}\right) \\
&= -\frac{\partial^2 b}{\partial\beta\partial G_l}y - \left(\frac{\partial^2 b}{\partial M_l^2}y - \frac{d^2c}{dM_l^2}\right)\left(-\frac{m_2}{m_1}\right)\left(-\frac{m_3}{m_1}\right) - \left(\frac{\partial b}{\partial M_l}y - \frac{dc}{dM_l}\right)\left(\frac{-m_1m_{23}+m_2m_{13}}{m_1^2}\right) \\
&> 0,
\end{aligned}
$$

because $\frac{\partial^2 b}{\partial\beta\partial G_l} \geq 0$, $\frac{\partial^2 b}{\partial M_l^2}y - \frac{d^2c}{dM_l^2} < 0$, $\frac{\partial b}{\partial M_l}y - \frac{dc}{dM_l} < 0$, and $-m_1m_{23}+m_2m_{13} \geq 0$. In addition,

$$\begin{aligned}
&\frac{db^*}{dG_r d\beta}y + \frac{d^2 c_r^*}{dG_r d\beta} \\
&= \frac{\partial^2 b}{\partial\beta\partial G_r}y + \frac{\partial^2 b}{\partial M_r^2}\left(-\frac{m_2}{m_1}\right)\frac{m_3}{m_1}y + \frac{\partial b}{\partial M_r}\left(-\frac{-m_1 m_{23} + m_2 m_{13}}{m_1^2}\right)y \\
&+ \frac{d^2 c}{dM_r^2}\left(\frac{m_2}{m_1}\right)\left(\frac{-m_3}{m_1}\right) + \frac{dc}{dM_r}\frac{m_1 m_{23} - m_2 m_{13}}{m_1^2} \\
&= \frac{\partial^2 b}{\partial\beta\partial G_r}y + \left(-\frac{\partial^2 b}{\partial M_r^2}y - \frac{d^2 c}{dM_r^2}\right)\left(-\frac{m_2}{m_1}\right)\left(-\frac{m_3}{m_1}\right) + \left(-\frac{\partial b}{\partial M_r}y - \frac{dc}{dM_r}\right)\left(\frac{-m_1 m_{23} + m_2 m_{13}}{m_1^2}\right) \\
&< 0,
\end{aligned}$$

because $\frac{\partial^2 b}{\partial\beta\partial G_r} \geq 0$, $-\frac{\partial^2 b}{\partial M_r^2}y - \frac{d^2 c}{dM_r^2} < 0$,$-\frac{\partial b}{\partial M_r}y - \frac{dc}{dM_r} < 0$, and $-m_1 m_{23} + m_2 m_{13} \geq 0$.

Next, we consider the first stage FOCs,

$$\begin{bmatrix} \frac{d^2 b^*}{dG_l^2}y - \frac{d^2 c^*}{dG_l^2} & \frac{d^2 b^*}{dG_r dG_l}y - \frac{\partial^2 c^*}{dG_r dG_l} \\ -\frac{d^2 b^*}{dG_r dG_l}y - \frac{d^2 c^*}{dG_r dG_l} & -\frac{d^2 b^*}{dG_r^2}y - \frac{d^2 c^*}{dG_r^2} \end{bmatrix}\begin{bmatrix} dG_l \\ dG_r \end{bmatrix} = \begin{bmatrix} -\frac{db^*}{dG_l d\beta}y + \frac{d^2 c^*}{dG_l d\beta} \\ \frac{db^*}{dG_r d\beta}y + \frac{d^2 c^*}{dG_r d\beta} \end{bmatrix}.d\beta$$

When $G_l = 1 - G_r$ and $M_l = M_r$, $\frac{\partial^2 c^*}{dG_r dG_l} = \frac{d^2 c^*}{dG_r dG_l} = 0$ so that

$$\begin{bmatrix} \frac{d^2 b^*}{dG_l^2}y - \frac{d^2 c^*}{dG_l^2} & \frac{d^2 b^*}{dG_r dG_l}y \\ -\frac{d^2 b^*}{dG_r dG_l}y & -\frac{d^2 b^*}{dG_r^2}y - \frac{d^2 c^*}{dG_r^2} \end{bmatrix}\begin{bmatrix} dG_l \\ dG_r \end{bmatrix} = \begin{bmatrix} -\frac{db^*}{dG_l d\beta}y + \frac{d^2 c^*}{dG_l d\beta} \\ \frac{db^*}{dG_r d\beta}y + \frac{d^2 c^*}{dG_r d\beta} \end{bmatrix} d\beta.$$

Note that $\frac{d^2 b^*}{dG_l^2}y - \frac{d^2 c^*}{dG_l^2} = -\frac{d^2 b^*}{dG_r^2}y - \frac{d^2 c^*}{dG_r^2}$. Hence,

$$\begin{aligned}
\frac{dG_l}{d\beta} &= \frac{\left(-\frac{db^*}{dG_l d\beta}y + \frac{d^2 c_l^*}{dG_l d\beta}\right)\left(-\frac{d^2 b^*}{dG_r^2}y - \frac{d^2 c^*}{dG_r^2}\right) - \left(\frac{db^*}{dG_r d\beta}y + \frac{d^2 c^*}{dG_r d\beta}\right)\frac{d^2 b^*}{dG_r dG_l}y}{\left(\frac{d^2 b^*}{dG_l^2}y - \frac{d^2 c_l^*}{dG_l^2}\right)^2 + \frac{d^2 b^*}{dG_r dG_l}y} \\
&= \frac{\left(-\frac{db^*}{dG_l d\beta}y + \frac{d^2 c_l^*}{dG_l d\beta}\right)\left(-\frac{d^2 b^*}{dG_r^2}y - \frac{d^2 c^*}{dG_r^2} + \frac{d^2 b^*}{dG_r dG_l}y\right)}{\left(\frac{d^2 b^*}{dG_l^2}y - \frac{d^2 c_l^*}{dG_l^2}\right)^2 + \left(\frac{d^2 b^*}{dG_r dG_l}y\right)^2} < 0,
\end{aligned}$$

because $-\frac{db^*}{dG_l d\beta}y + \frac{d^2 c_l^*}{dG_l d\beta} > 0$, $-\frac{d^2 b^*}{dG_r^2} + \frac{d^2 b^*}{dG_r dG_l} < 0$, and $\frac{d^2 c^*}{dG_r^2} > 0$. Furthermore,

$$\begin{aligned}
\frac{dG_r}{d\beta} &= \frac{\left(\frac{db^*}{dG_r d\beta}y + \frac{d^2 c_r^*}{dG_r d\beta}\right)\left(\frac{d^2 b^*}{dG_l^2}y - \frac{d^2 c^*}{dG_l^2}\right) + \left(-\frac{db^*}{dG_l d\beta}y + \frac{d^2 c^*}{dG_l d\beta}\right)\frac{d^2 b^*}{dG_r dG_l}y}{\left(\frac{d^2 b^*}{dG_l^2}y - \frac{d^2 c_l^*}{dG_l^2}\right)^2 + \frac{d^2 b^*}{dG_r dG_l}y} \\
&= \frac{\left(\frac{db^*}{dG_r d\beta}y + \frac{d^2 c_r^*}{dG_r d\beta}\right)\left(\frac{d^2 b^*}{dG_l^2}y - \frac{d^2 c^*}{dG_l^2} - \frac{d^2 b^*}{dG_r dG_l}y\right)}{\left(\frac{d^2 b^*}{dG_l^2}y - \frac{d^2 c_l^*}{dG_l^2}\right)^2 + \frac{d^2 b^*}{dG_r dG_l}y} > 0,
\end{aligned}$$

because $\frac{db^*}{dG_r d\beta}y + \frac{d^2 c_r^*}{dG_r d\beta} < 0$, $\frac{d^2 b^*}{dG_l^2} - \frac{d^2 b^*}{dG_r dG_l} < 0$, and $\frac{d^2 c^*}{dG_l^2} > 0$. □

We are now ready to prove Proposition 3.

Proof. First, consider the case of a one-sided threat. Suppose that, contrary to Proposition 3,

$$V_e^* < V_l^* + V_r^* \ ,$$

then regime e can mimic the choices of regime l, set $G_e = G_l^*$ and $M_e = M_l^*$, and obtain a payoff that is weakly greater than the sum of the net tax revenues of regimes l and r, which is a contradiction. Hence, it must be the case that

$$V_e^* \geq V_l^* + V_r^* \ .$$

In fact, the inequality has to be strict since regime r makes a non-zero military investment. Hence,

$$V_e^* > V_l^* + V_r^* \ .$$

Note that both V_e^* and $V_l^* + V_r^*$ are decreasing in Λ. By monotonicity, there exists a threshold $\Lambda = \dot{\Lambda}$ where $V_e^* > 0$ and $V_l^* + V_r^* = 0$; there exists another threshold $\Lambda = \ddot{\Lambda}(> \dot{\Lambda})$ where $V_e^* = 0$ and $V_l^* + V_r^* < 0$. By continuity, for all $\Lambda \in (\dot{\Lambda}, \ddot{\Lambda}]$, regime e is viable but regimes l and r are not (Implication 4).

Next, consider the case of a two-sided threat. For a centralized regime, when $\Lambda > \Lambda_{II}$,

$$\lambda\left(\Lambda, \frac{1}{2}(1-\delta)\right) = m\left(M_e^*, \delta/2, \beta\right),$$

so that

$$\frac{dM_e}{d\beta} = -\frac{m_3}{m_1} > 0.$$

Therefore,

$$\frac{dV_e}{d\beta} = -\frac{\partial c}{\partial M_e}\frac{dM_e}{d\beta} < 0.$$

Since $\frac{d^2 M_e}{d\beta^2} = -\frac{m_1 m_{33} - m_3 m_{13}}{(m_1)^2} \geq 0$,

$$\frac{d^2 V_e}{d\beta^2} = -\frac{\partial^2 c}{\partial M_e^2}\left(\frac{dM_e}{d\beta}\right)^2 - \frac{\partial c}{\partial M_e}\frac{d^2 M_e}{d\beta^2} < 0.$$

Moving on to interstate competition. Consider a symmetric interior equilibrium ($G_l^* \neq 0$ and $G_r^* \neq 1$). Given the symmetry, it suffices for us to focus on regime l alone. Recall that when the revolution constraint is binding,

$$\frac{\partial M_l}{\partial \beta} = \frac{\partial M_r}{\partial \beta} \geq 0 \text{ and } \frac{\partial G_l}{\partial \beta} = -\frac{\partial G_r}{\partial \beta} \leq 0,$$

and equality holds when $G_l^* = \frac{1}{2}(1-\delta)$ and $G_r^* = \frac{1}{2}(1+\delta)$. When β is large enough such that $G_l^* = \frac{1}{2}(1-\delta)$ and $G_r^* = \frac{1}{2}(1+\delta)$, $m\left(M_l, G_l^* - \frac{1}{2}(1-\delta), \beta\right) = \lambda\left(\Lambda, \frac{1}{2}(1-\delta)\right)$ and $m\left(M_r, (1-G_r^*) - \frac{1}{2}(1-\delta), \beta\right) = \lambda\left(\Lambda, \frac{1}{2}(1-\delta)\right)$. It follows that $\frac{\partial V_l}{\partial \beta} = 0$ because $\frac{\partial M_l}{\partial \beta} = \frac{\partial M_r}{\partial \beta} = 0$ and $\frac{\partial G_l}{\partial \beta} = -\frac{\partial G_r}{\partial \beta} = 0$. Since $\frac{dV_e}{d\beta} < 0$ and $\frac{d^2 V_e}{d\beta^2} < 0$, we conclude that when β is sufficiently large, $V_e^* < V_l^* + V_r^*$. □

A.2.4 More on Proposition 3

Part 2 of Proposition 3 states that under a two-sided threat, when $\Lambda \geq \Lambda_{II}$ and β is sufficiently large, $V_e^* < V_l^* + V_r^*$. The proof above (Section A.2.3) shows that as β increases to the level where the equilibrium capital city locations become $G_l^* = \frac{1}{2}(1-\delta)$ and $G_r^* = \frac{1}{2}(1+\delta)$, further increasing β will lead to an accelerating fall in V_e^* while $V_l^* + V_r^*$ remains constant. At some point, $V_e^* < V_l^* + V_r^*$ has to occur.

So far we have imposed minimal assumptions on the shapes of $m(\cdot)$ and $b(\cdot)$. Here, we show that the "sufficiently large" value of β that characterizes Proposition 3 can be much lower than the level defined above if we are willing to impose additional assumptions on the shapes of $m(\cdot)$ and $b(\cdot)$.

Assumption 3. *The military effectiveness function $m(M, t, \beta)$ satisfies $m_1(m_3 m_{22} - m_2 m_{23}) + m_2(m_2 m_{13} - m_3 m_{12}) \geq 0$, and the border function satisfies $\frac{\partial^2 b}{\partial G_l \partial G_r} = 0$.*

This assumption is somewhat technical as it imposes restrictions on cross-partial derivatives. Although not intuitive, it is not overly restrictive and is fulfilled by common specific functional forms. For example, suppose that $m(M, t, \beta) = g(\beta) f(M, t)$ where $g' < 0$ and $f_1 > 0$, $f_2 < 0$, $f_{22} < 0$ and

$f_{12} < 0$. Then,

$$\begin{aligned}
& m_1 \left(m_3 m_{22} - m_2 m_{23}\right) + m_2 \left(m_2 m_{13} - m_3 m_{12}\right) \\
& = g f_1 \left(g g' f f_{22} - g g' \left(f_2\right)^2\right) + g f_2 \left(g g' f_2 f_1 - g g' f f_{12}\right) \\
& = g^2 g' f \left(f_1 f_{22} - f_2 f_{12}\right) \\
& = g^2 \left(-g'\right) f \left(f_2 f_{12} - f_1 f_{22}\right) \geq 0.
\end{aligned}$$

As another example, suppose $m\left(M, t, \beta\right) = h\left(M\right) + k\left(\beta\right) \cdot f\left(t\right)$ where $h' > 0$, $k < 0$ $k' < 0$, $f' > 0$, and f is log-convex,

$$\begin{aligned}
& m_1 \left(m_3 m_{22} - m_2 m_{23}\right) + m_2 \left(m_2 m_{13} - m_3 m_{12}\right) \\
& = h' \left(\left(k' f\right)\left(k f''\right) - k k' \left(f'\right)^2\right) + k f'\left(0\right) \\
& = h' k k' \left(f f'' - \left(f'\right)^2\right) \geq 0.
\end{aligned}$$

An example of $b(\cdot)$ that satisfies Assumption 3 is

$$b\left(M_l, M_r, G_l, G_r, \beta\right) = \frac{G_r + G_l}{2} + \left(1 - \omega\right) \frac{M_l - M_r}{2\beta\left(G_r - G_l\right)} + \omega\left(\sqrt{G_l} - \sqrt{1 - G_r}\right),$$

where $\omega \in [0, 1]$.

We shall rework the proof of Proposition 3 by utilizing Assumption 3. Before doing so, it is useful to characterize the outcome of interstate competition with external threats.

Lemma 3. *For a two-sided threat where $\Lambda > \Lambda_{II}$, if β is sufficiently large and Assumption 3 holds, in the symmetric equilibrium,*

$$\begin{aligned}
& \left(\frac{dM_l}{dG_l} + \frac{dM_r}{dG_l}\right) \frac{\partial G_l}{\partial \beta} + \frac{\partial M_l}{\partial \beta} \geq 0, \text{ and} \\
& \left(\frac{dM_l}{dG_l} + \frac{dM_r}{dG_l}\right) \frac{\partial G_l}{\partial \beta} + \frac{\partial M_l}{\partial \beta} \geq 0,
\end{aligned}$$

where equality holds when $G_l^* = \frac{1}{2}\left(1 - \delta\right)$ and $G_r^* = \frac{1}{2}\left(1 + \delta\right)$.

Proof. If $\Lambda > \Lambda_{II}$ and β is sufficiently large, $\Lambda > \Lambda_{III}(\beta)$ because Λ_{III} is strictly decreasing in β. Note that when $\frac{d^2 b^*}{dG_r dG_l} = 0$,

$$\frac{dG_l}{d\beta} = \frac{-\frac{db^*}{dG_l d\beta} y + \frac{d^2 c_l^*}{dG_l d\beta}}{\frac{d^2 b^*}{dG_l^2} y - \frac{d^2 c_l^*}{dG_l^2}} < 0, \text{ and } \frac{dG_r}{d\beta} = \frac{\frac{db^*}{dG_r d\beta} y + \frac{d^2 c_r^*}{dG_r d\beta}}{-\frac{d^2 b^*}{dG_r^2} y - \frac{d^2 c_r^*}{dG_r^2}} > 0,$$

because

$$
\begin{aligned}
&\frac{\partial^2 b^*}{\partial G_l^2}y - \frac{d^2 c_l^*}{dG_l^2} \\
&= \frac{\partial^2 b}{\partial G_l^2}y + \frac{\partial^2 b}{\partial M_l^2}\left(-\frac{m_2}{m_1}\right)^2 y - \frac{d^2 c}{dM_l^2}\left(-\frac{m_2}{m_1}\right)^2 + \left(\frac{\partial b}{\partial M_l}y - \frac{dc}{dM_l}\right)\left(\frac{-m_1 m_{22} + m_2 m_{12}}{(m_1)^2}\right) \\
&= \frac{\partial^2 b}{\partial G_l^2}y + \left(\frac{\partial^2 b}{\partial M_l^2}y - \frac{d^2 c}{dM_l^2}\right)\left(-\frac{m_2}{m_1}\right)^2 + \left(\frac{\partial b}{\partial M_l}y - \frac{dc}{dM_l}\right)\left(\frac{-m_1 m_{22} + m_2 m_{12}}{(m_1)^2}\right) \\
&< 0,
\end{aligned}
$$

and

$$
\begin{aligned}
&-\frac{\partial^2 b^*}{\partial G_r^2}y - \frac{d^2 c_l^*}{dG_l^2} \\
&= -\frac{\partial^2 b}{\partial G_r^2}y - \frac{\partial^2 b}{\partial M_r^2}\left(-\frac{m_2}{m_1}\right)^2 y - \frac{d^2 c}{dM_r^2}\left(-\frac{m_2}{m_1}\right)^2 + \left(-\frac{\partial b}{\partial M_r}y - \frac{dc}{dM_r}\right)\left(\frac{-m_1 m_{22} + m_2 m_{12}}{(m_1)^2}\right) \\
&= -\frac{\partial^2 b}{\partial G_r^2}y + \left(-\frac{\partial^2 b}{\partial M_r^2}y - \frac{d^2 c}{dM_r^2}\right)\left(-\frac{m_2}{m_1}\right)^2 + \left(-\frac{\partial b}{\partial M_r}y - \frac{dc}{dM_r}\right)\left(\frac{-m_1 m_{22} + m_2 m_{12}}{(m_1)^2}\right) \\
&< 0.
\end{aligned}
$$

Now,

$$
\begin{aligned}
&\left(\frac{dM_l}{dG_l} + \frac{dM_r}{dG_l}\right)\frac{dG_l}{d\beta} + \frac{dM_l}{d\beta} \\
&= \left(-\frac{m_2}{m_1}\right)\frac{-\frac{\partial^2 b}{\partial\beta\partial G_l}y - \left(\frac{\partial^2 b}{\partial M_l^2}y - \frac{d^2 c}{dM_l^2}\right)\left(-\frac{m_2}{m_1}\right)\left(-\frac{m_3}{m_1}\right) - \left(\frac{\partial b}{\partial M_l}y - \frac{dc}{dM_l}\right)\left(\frac{-m_1 m_{23} + m_2 m_{13}}{m_1^2}\right)}{\frac{\partial^2 b}{\partial G_l^2}y + \left(\frac{\partial^2 b}{\partial M_l^2}y - \frac{d^2 c}{dM_l^2}\right)\left(-\frac{m_2}{m_1}\right)^2 + \left(\frac{\partial b}{\partial M_l}y - \frac{dc}{dM_l}\right)\left(\frac{-m_1 m_{22} + m_2 m_{12}}{(m_1)^2}\right)} - \frac{m_3}{m_1} \\
&= \frac{\frac{\partial^2 b}{\partial G_l^2}\left(-\frac{m_3}{m_1}\right)y + \frac{\partial^2 b}{\partial\beta\partial G_l}\left(-\frac{m_2}{m_1}\right)y + \left(\frac{\partial b}{\partial M_l}y - \frac{dc}{dM_l}\right)\left(\left(-\frac{m_3}{m_1}\right)\frac{-m_1 m_{22} + m_2 m_{12}}{(m_1)^2} - \left(-\frac{m_2}{m_1}\right)\frac{-m_1 m_{23} + m_2 m_{13}}{m_1^2}\right)}{\frac{\partial^2 b}{\partial G_l^2}y + \left(\frac{\partial^2 b}{\partial M_l^2}y - \frac{d^2 c}{dM_l^2}\right)\left(-\frac{m_2}{m_1}\right)^2 + \left(\frac{\partial b}{\partial M_l}y - \frac{dc}{dM_l}\right)\left(\frac{-m_1 m_{22} + m_2 m_{12}}{(m_1)^2}\right)} \\
&= \frac{\frac{\partial^2 b}{\partial G_l^2}\left(-\frac{m_3}{m_1}\right)y + \frac{\partial^2 b}{\partial\beta\partial G_l}\left(-\frac{m_2}{m_1}\right)y + \left(\frac{\partial b}{\partial M_l}y - \frac{dc}{dM_l}\right)\left(\frac{m_1(m_3 m_{22} - m_2 m_{23}) + m_2(m_2 m_{13} - m_3 m_{12})}{(m_1)^3}\right)}{\frac{\partial^2 b}{\partial G_l^2}y + \left(\frac{\partial^2 b}{\partial M_l^2}y - \frac{d^2 c}{dM_l^2}\right)\left(-\frac{m_2}{m_1}\right)^2 + \left(\frac{\partial b}{\partial M_l}y - \frac{dc}{dM_l}\right)\left(\frac{-m_1 m_{22} + m_2 m_{12}}{(m_1)^2}\right)} > 0
\end{aligned}
$$

because $\frac{\partial^2 b}{\partial\beta\partial G_l}\left(-\frac{m_2}{m_1}\right) \leq 0$, $\frac{\partial^2 b}{\partial G_l^2}\left(-\frac{m_3}{m_1}\right) \leq 0$, $\frac{\partial b}{\partial M_l}y - \frac{dc}{dM_l} < 0$, and $m_1(m_3 m_{22} - m_2 m_{23}) + m_2(m_2 m_{13} - m_3 m_{12}) \geq 0$.

Next,

$$
\begin{aligned}
&\left(\frac{dM_l}{dG_r}+\frac{dM_r}{dG_r}\right)\frac{dG_r}{d\beta}+\frac{dM_r}{d\beta}\\
&=\left(\frac{m_2}{m_1}\right)\frac{\frac{\partial^2 b}{\partial\beta\partial G_r}y+\left(-\frac{\partial^2 b}{\partial M_r^2}y-\frac{d^2c}{dM_r^2}\right)\left(-\frac{m_2}{m_1}\right)\left(-\frac{m_3}{m_1}\right)+\left(-\frac{\partial b}{\partial M_r}y-\frac{dc}{dM_r}\right)\left(\frac{-m_1m_{23}+m_2m_{13}}{m_1^2}\right)}{-\frac{\partial^2 b}{\partial G_r^2}y+\left(-\frac{\partial^2 b}{\partial M_r^2}y-\frac{d^2c}{dM_r^2}\right)\left(-\frac{m_2}{m_1}\right)^2+\left(-\frac{\partial b}{\partial M_r}y-\frac{dc}{dM_r}\right)\left(\frac{-m_1m_{22}+m_2m_{12}}{(m_1)^2}\right)}-\frac{m_3}{m_1}\\
&=\frac{\frac{\partial^2 b}{\partial\beta\partial G_r}\frac{m_2}{m_1}y+\left(-\frac{\partial^2 b}{\partial G_r^2}\right)\left(-\frac{m_3}{m_1}\right)y+\left(-\frac{\partial b}{\partial M_r}y-\frac{dc}{dM_r}\right)\left(\left(-\frac{m_3}{m_1}\right)\frac{-m_1m_{22}+m_2m_{12}}{(m_1)^2}+\frac{m_2}{m_1}\frac{-m_1m_{23}+m_2m_{13}}{m_1^2}\right)}{-\frac{\partial^2 b}{\partial G_r^2}y+\left(-\frac{\partial^2 b}{\partial M_r^2}y-\frac{d^2c}{dM_r^2}\right)\left(-\frac{m_2}{m_1}\right)^2+\left(-\frac{\partial b}{\partial M_r}y-\frac{dc}{dM_r}\right)\left(\frac{-m_1m_{22}+m_2m_{12}}{(m_1)^2}\right)}\\
&=\frac{\frac{\partial^2 b}{\partial\beta\partial G_r}\frac{m_2}{m_1}y+\left(-\frac{\partial^2 b}{\partial G_r^2}\right)\left(-\frac{m_3}{m_1}\right)y+\left(-\frac{\partial b}{\partial M_r}y-\frac{dc}{dM_r}\right)\left(\frac{m_1(m_3m_{22}-m_2m_{23})+m_2(m_2m_{13}-m_3m_{12})}{(m_1)^3}\right)}{-\frac{\partial^2 b}{\partial G_r^2}y+\left(-\frac{\partial^2 b}{\partial M_r^2}y-\frac{d^2c}{dM_r^2}\right)\left(-\frac{m_2}{m_1}\right)^2+\left(-\frac{\partial b}{\partial M_r}y-\frac{dc}{dM_r}\right)\left(\frac{-m_1m_{22}+m_2m_{12}}{(m_1)^2}\right)}\geq 0
\end{aligned}
$$

because $\frac{\partial^2 b}{\partial\beta\partial G_r}\frac{m_2}{m_1}\leq 0$, $\left(-\frac{\partial^2 b}{\partial G_r^2}\right)\left(-\frac{m_3}{m_1}\right)\leq 0$, $\left(-\frac{\partial b}{\partial M_r}y-\frac{dc}{dM_r}\right)<0$, and $m_1(m_3m_{22}-m_2m_{23})+m_2(m_2m_{13}-m_3m_{12})\geq 0$. $\square$

We are ready to prove Proposition 3 (again).

Proof. Consider $\Lambda\geq\Lambda_{III}(\beta)$ such that the revolution constraint is binding. By envelope theorem,

$$
\begin{aligned}
\frac{dV_l^*}{d\beta}&=\left(\left(\frac{\partial b}{\partial G_r}+\frac{\partial b}{\partial M_l}\frac{\partial M_l}{\partial G_r}+\frac{\partial b}{\partial M_r}\frac{\partial M_r}{\partial G_r}\right)y-\frac{\partial c}{\partial M_l}\frac{\partial M_l}{\partial G_r}\right)\frac{\partial G_r}{\partial\beta}\\
&\quad+\left(\frac{\partial b}{\partial M_l}y-\frac{\partial c}{\partial M_l}\right)\frac{\partial M_l}{\partial\beta}+\frac{\partial b}{\partial M_r}\frac{\partial M_r}{\partial\beta}y\\
&=\left(\left(\frac{\partial b}{\partial G_r}+\frac{\partial b}{\partial M_l}\frac{\partial M_l}{\partial G_r}+\frac{\partial b}{\partial M_r}\frac{\partial M_r}{\partial G_r}\right)y-\frac{\partial c}{\partial M_l}\frac{\partial M_l}{\partial G_r}\right)\frac{\partial G_r}{\partial\beta}-\frac{\partial c}{\partial M_r}\frac{\partial M_r}{\partial\beta}\\
&=\left(-\frac{\partial c}{\partial M_r}\frac{\partial M_r}{\partial G_r}+-\frac{\partial c}{\partial M_l}\frac{\partial M_l}{\partial G_r}\right)\frac{\partial G_r}{\partial\beta}-\frac{\partial c}{\partial M_r}\frac{\partial M_r}{\partial\beta}.
\end{aligned}
$$

Since $\frac{\partial c}{\partial M_l}=\frac{\partial c}{\partial M_r}$ in the symmetric equilibrium,

$$
\frac{dV_l^*}{d\beta}=-\frac{\partial c}{\partial M_r}\left(\left(\frac{\partial M_r}{\partial G_r}+\frac{\partial M_l}{\partial G_r}\right)\frac{\partial G_r}{\partial\beta}+\frac{\partial M_r}{\partial\beta}\right)\leq 0.
$$

Hence, if $\Lambda\geq\Lambda_{II}$, as β increases, at some point the revolution constraint will be binding. From this point onward, further increasing β initially leads to a continuous fall in $V_l^*+V_r^*$ but once the capital cities settle at $G_l^*=\frac{1}{2}(1-\delta)$ and $G_r^*=\frac{1}{2}(1+\delta)$, the fall stops and $V_l^*+V_r^*$ remains constant as β increases further. Since $\frac{dV_e^*}{d\beta}<0$ and $\frac{d^2V_e^*}{d\beta^2}<0$, there exists $\hat{\beta}$ such that for all $\beta>\hat{\beta}$,

$V_l^* + V_r^* > V_e^*$ and $G_l^* \geq \frac{1}{2}(1-\delta)$ and $G_r^* \leq \frac{1}{2}(1+\delta)$. Note that we no longer require β to reach the level that realizes $G_l^* = \frac{1}{2}(1-\delta)$ and $G_r^* = \frac{1}{2}(1+\delta)$ for $V_l^* + V_r^* > V_e^*$ to hold. $\square$

A.2.5 Proof of Positive Tax Reimbursement when $S = 1$

In Section 3.5, we claim that if the cost function of military investment is sufficiently convex, the empire will provide some tax reimbursement instead of relying solely on building the military to satisfy the revolution constraint. Here, we provide the proof.

Claim 4. *If $\frac{\partial c(M)}{\partial M} > \frac{\partial m(M,0,\beta)}{\partial M}$ for all M, regime e provides a strictly positive tax reimbursement ($R_e > 0$).*

Proof. Suppose regime e faces a one-sided threat (The proof for a two-sided threat is similar). The optimization problem is given by:

$$\max_{G_e,M_e,R_e} V_e = y - c\left(M_e\right) - R_e$$

subject to

$$\begin{aligned}
&R_e \geq 0, \\
&m\left(M_e, G_e - (1-\delta), \beta\right) + R_e = \lambda\left(\Lambda, 1-\delta\right), \text{ and} \\
&m\left(M_e, \left|x - G_e\right|, \beta\right) \geq \lambda\left(\Lambda, x\right) \text{ for some } x \in \left[0, x^*\left(\Lambda\right)\right].
\end{aligned}$$

Let $\bar{x} \in \left[0, x^*\left(\Lambda\right)\right]$ such that $m(M_e, \left|\bar{x} - G_e\right|, \beta) = \lambda\left(\Lambda, \bar{x}\right)$. Since $\lambda_2 < 0$, $\bar{x} \geq G_e$ (otherwise, the empire can increase its net tax revenue by increasing G_e and decreasing M_e or R_e). If $R_e = 0$, then $G_e = 1 - \delta$ since $m_2 < 0$. If $G_e^* > 1 - \delta$, it must be the case that $R_e^* > 0$ (otherwise, if $R_e^* = 0$, the empire can increase its net tax revenue by decreasing G_e and M_e simultaneously). Therefore, it suffices to compare $R_e^* = 0$ and $R_e^* > 0$ when $G_e = 1 - \delta$. When $G_e = 1 - \delta$, the Lagrangian optimization problem is:

$$\max_{M_e,R_e} \mathcal{L}_e = y - c\left(M_e\right) - R_e + \phi R_e + \gamma\left(m\left(M_e, 0, \beta\right) + R_e - \lambda\left(\Lambda, 1-\delta\right)\right)$$

where ϕ and γ are the Lagrangian multipliers. The first order conditions are given by:

$$\begin{aligned}
M_e: &\quad - c_M + \gamma m_1\left(M_e, 0, \beta\right) = 0, \\
R_e: &\quad - 1 + \phi + \gamma = 0, \\
\phi: &\quad \phi R_e \geq 0 \text{ and either } \phi = 0 \text{ or } R_e = 0, \\
\gamma: &\quad \lambda\left(\Lambda, 1-\delta\right) = m\left(M_e, 0, \beta\right) + R_e.
\end{aligned}$$

Suppose $R_e^* = 0$. Then,

$$\begin{aligned} c_M\left(M_e^*\right) &= \left(1-\phi^*\right) m_1\left(M_e^*, 0, \beta\right), \text{ and} \\ \lambda\left(\Lambda, 1-\delta\right) &= m\left(M_e^*, 0, \beta\right). \end{aligned}$$

Note that $dM_e^*/d\beta = 0$ since $dm\left(M_e^*, 0, \beta\right)/d\beta = 0$.

Alternatively, suppose $R_e^{**} > 0$. This implies $\phi^{**} = 0$, and

$$\begin{aligned} c_M\left(M_e^{**}\right) &= m_1\left(M_e^{**}, 0, \beta\right), \text{ and} \\ \lambda\left(\Lambda, 1-\delta\right) &= m\left(M_e^{**}, 0, \beta\right) + R_e^{**}. \end{aligned}$$

Note that $dM_e^{**}/d\beta = 0$ and $dR_e^{**}/d\beta = 0$ since $dm_1\left(M_e^{**}, 0, \beta\right)/d\beta = 0$. Furthermore,

$$\begin{aligned} m\left(M_e^*, 0, \beta\right) &= m\left(M_e^{**}, 0, \beta\right) + R_e^{**} \\ &> m\left(M_e^{**}, 0, \beta\right). \end{aligned}$$

This implies that $M_e^* > M_e^{**}$. Now, let

$$\begin{aligned} \Psi &\equiv c\left(M_e^*\right) - c\left(M_e^{**}\right) - R_e^{**} \\ &= c\left(M_e^*\right) - c\left(M_e^{**}\right) - \left(m\left(M_e^*, 0, \beta\right) - m\left(M_e^{**}, 0, \beta\right)\right) \\ &= \left(c\left(M_e^*\right) - m\left(M_e^*, 0, \beta\right)\right) - \left(c\left(M_e^{**}\right) - m\left(M_e^{**}, 0, \beta\right)\right) \end{aligned}$$

The empire should set $R_e > 0$ only if $\Psi > 0$. Since $M_e^* > M_e^{**}$, $\Psi > 0$ if for all $M \geq 0$,

$$\frac{\partial c\left(M\right)}{\partial M} > \frac{\partial m\left(M, 0, \beta\right)}{\partial M}.$$

□

A.3 Multiple Regimes ($S \geq 3$)

In this section, we extend the model from $S \leq 2$ to $S \geq 3$.

Consider the case where there are $S \geq 3$ regimes in the continent and regime s is to the left of regime $s+1$ for all $s < S$. In this two-stage game, all regimes first simultaneously choose the location of their capital cities $G_s \in [0,1]$ where $G_1 < G_2 < \cdots < G_S$. Let $b_{s,s+1}$ denote the border between regime s and regime $s+1$. We restate Assumption 2 as follows:

Assumption' 2. *The border $b_{s,s+1}(G_s, G_{s+1}, M_s, M_{s+1}, \beta) \in [G_s, G_{s+1}]$ satisfies the following properties: for all $s = 1, \ldots, S-1$,*

(a) (Monotonicity) $\frac{\partial b_{s,s+1}}{\partial M_s} > 0$, $\frac{\partial b_{s,s+1}}{\partial M_{s+1}} < 0$; $\frac{\partial b_{s,s+1}}{\partial G_s} > 0$, $\frac{\partial b_{s,s+1}}{\partial G_{s+1}} > 0$,

(b) (Convexity) $\frac{\partial^2 b_{s,s+1}}{\partial M_s^2} \leq 0$, $\frac{\partial^2 b_{s,s+1}}{\partial M_{s+1}^2} \leq 0$; $\frac{\partial^2 b_{s,s+1}}{\partial G_s^2} \leq 0$, $\frac{\partial^2 b_{s,s+1}}{\partial G_{s+1}^2} \geq 0$.

After observing the locations of the capital cities, the regimes simultaneously decide on their military investments $M_s \geq 0$. This is a complete information game and we employ subgame-perfect equilibrium as the solution concept.

Lemma' 1. *When there is no external threat, given the locations of capital cities $(G_1, \ldots, G_S)$, the equilibrium military investments $(M_1^*, \ldots, M_S^*)$ satisfy:*

$$\frac{\partial b_{1,2}}{\partial M_1} y - \frac{\partial c}{\partial M_1} = 0,$$

$$\left(\frac{\partial b_{s,s+1}}{\partial M_s} - \frac{\partial b_{s-1,s}}{\partial M_s} \right) y - \frac{\partial c}{\partial M_s} = 0, \text{ for all } s = 2, \ldots, S-1, \text{ and}$$

$$\frac{\partial b_{S-1,S}}{\partial M_S} y - \frac{\partial c}{\partial M_S} = 0.$$

For the equilibrium locations of capital cities $(G_1^*, \ldots, G_S^*)$ that satisfy $G_1^* \neq 0$ and $G_S^* \neq 1$ (interior solutions),

$$\left(\frac{\partial b_{1,2}}{\partial G_1} + \frac{\partial b_{1,2}}{\partial M_1} \frac{\partial M_1}{\partial G_1} + \frac{\partial b_{1,2}}{\partial M_2^*} \frac{\partial M_2}{\partial G_1} \right) y - \frac{\partial c}{\partial M_1} \frac{\partial M_1}{\partial G_1} = 0, \text{ and}$$

$$\left(\frac{\partial b_{s,s+1}}{\partial G_s} + \frac{\partial b_{s,s+1}}{\partial M_s^*} \frac{\partial M_s}{\partial G_s} + \frac{\partial b_{s,s+1}}{\partial M_{s+1}} \frac{\partial M_{s+1}}{\partial G_s} \right) y -$$
$$\left(\frac{\partial b_{s-1,s}}{\partial G_s} + \frac{\partial b_{s-1,s}}{\partial M_{s-1}} \frac{\partial M_{s-1}}{\partial G_s} + \frac{\partial b_{s-1,s}}{\partial M_s} \frac{\partial M_s}{\partial G_s} \right) y - \frac{\partial c}{\partial M_s^*} \frac{\partial M_s}{\partial G_s} = 0 \text{ for all } s = 2, \ldots, S-1,$$

$$-\left(\frac{\partial b_{S-1,S}}{\partial G_S} + \frac{\partial b_{S-1,S}}{\partial M_{S-1}} \frac{\partial M_{S-1}}{\partial G_S} + \frac{\partial b_{S-1,S}}{\partial M_S} \frac{\partial M_S}{\partial G_S} \right) y - \frac{\partial c}{\partial M_S} \frac{\partial M_S}{\partial G_S} = 0.$$

Proof. Consider the second stage of the game. Given $G_1, \dots, G_S$, the optimization problems for the regimes are

$$\max_{M_1} V_1 = b_{1,2}\left(G_1, G_2, M_1, M_2, \beta\right) y - c\left(M_1\right),$$
$$\max_{M_s} V_s = (b_{s,s+1}\left(G_s, G_{s+1}, M_s, M_{s+1}, \beta\right) - b_{s-1,s}\left(G_s, G_{s+1}, M_{s-1}, M_s, \beta\right))y - c\left(M_s\right) \text{ for } s = 2, \dots, S-1,$$
$$\max_{M_S} V_S = \left(1 - b_{S-1,S}\left(G_{S-1}, G_S, M_{S-1}, M_S, \beta\right)\right) y - c\left(M_S\right).$$

The respective FOCs are

$$\begin{aligned} \frac{\partial b_{1,2}}{\partial M_1} - \frac{\partial c}{\partial M_1} &= 0, \\ \left(\frac{\partial b_{s,s+1}}{\partial M_s} - \frac{\partial b_{s-1,s}}{\partial M_s}\right) y - \frac{\partial c}{\partial M_s} &= 0, \text{ for } s = 2, \dots, S-1, \text{and} \\ -\frac{\partial b_{S-1,S}}{\partial M_S} y - \frac{\partial c}{\partial M_S} &= 0. \end{aligned}$$

Since $\frac{\partial^2 c}{\partial^2 M_s} < 0$ for all s, the SOCs are guaranteed if $\frac{\partial^2 b_{s,s+1}}{\partial M_s^2} \leq 0$ and $\frac{\partial^2 b_{s,s+1}}{\partial M_{s+1}^2} \geq 0$ for all $s = 1, \dots, S-1$. Given some second-stage equilibrium military investments $M_1\left(G_1, G_2\right)$, $M_s\left(G_{s-1}, G_s, G_{s+1}\right)$ for all $s = 2, \dots, S-1$ and $M_S\left(G_{S-1}, G_S\right)$, consider the first stage when regimes choose their capitals:

$$\max_{G_1} V_1 = b_{1,2}\left(G_1, G_2, M_1, M_2, \beta\right) y - c\left(M_1\right),$$
$$\max_{G_s} V_s = (b_{s,s+1}\left(G_s, G_{s+1}, M_s, M_{s+1}, \beta\right) - b_{s-1,s}\left(G_s, G_{s+1}, M_{s-1}^*, M_s, \beta\right))y - c\left(M_s^*\right) \text{ for } s = 2, \dots, S-1\ ,$$
$$\max_{G_S} V_S = \left(1 - b_{S-1,S}\left(G_{S-1}, G_2, M_{S-1}, M_S, \beta\right)\right) y - c\left(M_S\right).$$

The respective FOCs are

$$\left(\frac{\partial b_{1,2}}{\partial G_1} + \frac{\partial b_{1,2}}{\partial M_1}\frac{\partial M_1}{\partial G_1} + \frac{\partial b_{1,2}}{\partial M_2^*}\frac{\partial M_2}{\partial G_1}\right) y - \frac{\partial c}{\partial M_1}\frac{\partial M_1}{\partial G_1} = 0,$$

$$\begin{aligned} &\left(\frac{\partial b_{s,s+1}}{\partial G_s} + \frac{\partial b_{s,s+1}}{\partial M_s^*}\frac{\partial M_s}{\partial G_s} + \frac{\partial b_{s,s+1}}{\partial M_{s+1}}\frac{\partial M_{s+1}}{\partial G_s}\right) y - \\ &\left(\frac{\partial b_{s-1,s}}{\partial G_s} + \frac{\partial b_{s-1,s}}{\partial M_{s-1}}\frac{\partial M_{s-1}}{\partial G_s} + \frac{\partial b_{s-1,s}}{\partial M_s}\frac{\partial M_s}{\partial G_s}\right) y - \frac{\partial c}{\partial M_s^*}\frac{\partial M_s}{\partial G_s} = 0 \text{ for all } s = 2, \dots, S-1, \end{aligned}$$

$$-\left(\frac{\partial b_{S-1,S}}{\partial G_S} + \frac{\partial b_{S-1,S}}{\partial M_{S-1}}\frac{\partial M_{S-1}}{\partial G_S} + \frac{\partial b_{S-1,S}}{\partial M_S}\frac{\partial M_S}{\partial G_S}\right) y - \frac{\partial c}{\partial M_S}\frac{\partial M_S}{\partial G_S} = 0.$$

□

Proposition' 2 (Interstate Competition)**.** *Let $\hat{\delta}$ denote the fraction of the continent that is protected from the external threat in equilibrium (i.e. $\hat{\delta} = \sum_{s=1}^{S} |\mathbb{D}_s|$). Consider the symmetric equilibrium where $M_s = M_{S-s+1}$ and $G_s = 1 - G_{S-s+1}$ for all $s = 1, \ldots, S$. When the threat is two-sided:*

1. *If $\Lambda \leq \Lambda_{III}(\beta)$, the revolution constraint does not bind. The equilibrium military investments and location of capitals are the same as in the case when $\Lambda = 0$.*

2. *If $\Lambda > \Lambda_{III}(\beta)$, the revolution constraint binds.*

Proof. Recall that $\Lambda_{III}(\beta)$ denotes the value of Λ that solves:

$$\lambda\left(\Lambda_{III}(\beta), \frac{1}{2}(1-\delta)\right) = m\left(M_1, G_1 - \frac{1}{2}(1-\delta), \beta\right).$$

It is clear that if $\Lambda \leq \Lambda_{III}(\beta)$, the revolution constraint does not bind, and vice versa.

$\square$

Proposition' 3. ***(Viability)***

1. *Under a one-side threat, $V_e^* > V_1^* + V_2^* + ... + V_S^*$.*

2. *Under a two-sided threat, if $\frac{\partial M_s}{\partial \beta} \leq 0$ for all $s \neq 1, S$, $\lim_{\beta \to \infty} \left|\frac{\partial M_1}{\partial \beta}\right| = \lim_{\beta \to \infty} \left|\frac{\partial M_S}{\partial \beta}\right| = 0$, and $\lim_{\beta \to \infty} \left|\frac{\partial G_s}{\partial \beta}\right| = 0$ for all $s = 1, \ldots, S$, then when β is sufficiently large, $V_e^* < V_1^* + V_2^* + ... + V_S^*$.*

Proof. The proof of the first part is similar to the two-regime case (See Appendix A.2.3). Now consider a two-sided threat. Recall that the case for centralization is $\frac{dV_e^*}{d\beta} < 0$ and $\frac{d^2 V_e^*}{d\beta^2} < 0$. It suffices to show that $\frac{d\sum_{s=1}^{S} V_s^*}{d\beta} \geq 0$. First note that

$$\begin{aligned}
\frac{d\sum_{s=1}^{S} V_s^*}{d\beta} = &- \frac{\partial c}{\partial M_1}\left(\frac{\partial M_1}{\partial G_1}\frac{\partial G_1}{\partial \beta} + \frac{\partial M_1}{\partial G_2}\frac{\partial G_2}{\partial \beta}\right) - \frac{\partial c}{\partial M_S}\left(\frac{\partial M_S}{\partial G_{S-1}}\frac{\partial G_{S-1}}{\partial \beta} + \frac{\partial M_S}{\partial G_S}\frac{\partial G_S}{\partial \beta}\right) \\
&- \sum_{s=1}^{S} \frac{\partial c}{\partial M_s}\left(\frac{\partial M_s}{\partial G_{s-1}}\frac{\partial G_{s-1}}{\partial \beta} + \frac{\partial M_s}{\partial G_s}\frac{\partial G_s}{\partial \beta} + \frac{\partial M_s}{\partial G_{s+1}}\frac{\partial G_{s+1}}{\partial \beta}\right) - \sum_{s=1}^{S} \frac{\partial c}{\partial M_s}\frac{\partial M_s}{\partial \beta}
\end{aligned}$$

If β is sufficiently large, $\frac{\partial G_s}{\partial \beta} \approx 0$ for $s = 1, \ldots, S$, $\frac{\partial M_1}{\partial \beta} \approx 0$, and $\frac{\partial M_S}{\partial \beta} \approx 0$. Therefore,

$$\frac{d\sum_{s=1}^{S} V_s^*}{d\beta} \approx -\sum_{s=2}^{S-1} \frac{\partial c}{\partial M_s}\frac{\partial M_s}{\partial \beta} \geq 0.$$

$\square$

Proposition' 3 imposes restrictions on the shapes of $m(.)$, $b(.)$, $c(.)$, and $\lambda(.)$ such that (1) regimes respond to a decrease in the military's power projection capability by decreasing military investment ($\frac{\partial M_s}{\partial \beta} \leq 0$);[45] and (2) they increasingly do not adjust their military investments and capital locations in response to increases in the value of β. In the next section (A.4), we provide a parametric example to illustrate the case when $S = 3$.

[45]This is somewhat analogous to requiring that military investment is not a "Giffen good" and regimes do not consume more as price rises.

A.4 Model with Imposed Functional Forms

In Section 3, we impose minimal restrictions on the shapes of $c(.)$, $m(.)$, $b(.)$, and $\lambda(.)$ and derive results based on the symmetric equilibrium. Here, we provide two functional form examples—for $S = 2$ and $S = 3$ respectively—to show that under plausible parametric assumptions, the symmetric equilibrium is, in fact, the unique equilibrium.

In what follows (A.4.1 and A.4.2), we assume that the military expenditure function takes the form $c\left(M\right) = \theta M^2$ where $\theta > 0$ is a cost parameter and the military effectiveness function takes the form $m\left(M, t\right) = M - \beta t^2$.

A.4.1 Two-Regime Continent

For $S = 2$, we define the border b as

$$b\left(M_l, M_r, G_l, G_r, \beta\right) = \frac{G_r + G_l}{2} + \left(1-\omega\right)\frac{M_l - M_r}{2\beta\left(G_r - G_l\right)} + \omega\left(\sqrt{G_l} - \sqrt{1 - G_r}\right),$$

where $\omega \in [0, 1]$. When $\omega = 0$,

$$b\left(M_l, M_r, G_l, G_r, \beta\right) = \frac{G_r + G_l}{2} + \frac{M_l - M_r}{2\beta\left(G_r - G_l\right)}.$$

The border is defined as the point between the capital cities where the regimes are equal in military strength if there exists such a location. In this case, b solves the following equation

$$m(M_l, b - G_l, \beta) = m(M_r, G_r - b, \beta).$$

When $\omega = 1$,

$$b\left(M_l, M_r, G_l, G_r, \beta\right) = \frac{G_r + G_l}{2} + \sqrt{G_l} - \sqrt{1 - G_r}.$$

In this case, military investments do not influence the location of the border. The border is simply the mid-point between the capital cities with adjustments for the uncontested region controlled by each regime. The following lemma shows the unique equilibrium is symmetric.

Lemma 4. *In the absence of external threats, the equilibrium is unique and symmetric. Regimes l and r invest the same amount on the military,*

$$M_l^* = M_r^* = \frac{y\left(1-\omega\right)}{4\beta\theta\left(G_r - G_l\right)},$$

and the capital city locations are symmetric, $G_l^* = 1 - G_r^* = G^*$, *where*

$$\frac{1}{2} + \frac{\omega}{2\sqrt{G^*}} - \frac{y(1-\omega)}{8\beta^2\theta(1-2G^*)^3} = 0.$$

Proof. We use backward induction and solve first the military investment problem given the capital locations before solving for the location of the two capitals.

$$\begin{aligned}\max_{M_l} V_l &= by - \theta M_l^2 \\ &= \left(\frac{G_r + G_l}{2} + (1-\omega)\frac{M_l - M_r}{2\beta(G_r - G_l)} + \omega\left(\sqrt{G_l} - \sqrt{1-G_r}\right)\right) y - \theta M_l^2.\end{aligned}$$

The FOC is

$$\frac{y(1-\omega)}{2\beta(G_r - G_l)} - 2\theta M_l = 0,$$

or

$$M_l^* = \frac{y(1-\omega)}{4\beta\theta(G_r - G_l)}.$$

It is easy to verify that the SOC is negative and

$$M_r^* = \frac{y(1-\omega)}{4\beta\theta(G_r - G_l)}.$$

Note that

$$\frac{\partial M_r^*}{\partial G_r} + \frac{\partial M_l^*}{\partial G_r} = -\frac{M_l^* + M_r^*}{G_r - G_l}.$$

In the first stage,

$$\begin{aligned}\max_{G_l} V_l &= \left(\frac{G_r + G_l}{2} + (1-\omega)\frac{M_l^* - M_r^*}{2\beta(G_r - G_l)} + \omega\left(\sqrt{G_l} - \sqrt{1-G_r}\right)\right) y - \theta M_l^{*2} \\ &= \left(\frac{G_r + G_l}{2} + \omega\left(\sqrt{G_l} - \sqrt{1-G_r}\right)\right) y - \theta\left(\frac{y(1-\omega)}{4\beta\theta(G_r - G_l)}\right)^2.\end{aligned}$$

Hence,

$$\frac{\partial V_l}{\partial G_l} = \left(\frac{1}{2} + \frac{\omega}{2\sqrt{G_l}}\right) y - \frac{(1-\omega)^2 y^2}{8\beta^2\theta(G_r - G_l)^3} = 0.$$

Similarly,

$$V_r = \left(1 - \left(\frac{G_r + G_l}{2} + \omega\left(\sqrt{G_l} - \sqrt{1-G_r}\right)\right)\right) y - \theta\left(\frac{(1-\omega) y}{4\beta\theta(G_r - G_l)}\right)^2,$$

so that

$$\frac{\partial V_r}{\partial G_r} = -\left(\frac{1}{2} + \frac{\omega}{2\sqrt{1-G_r}}\right) y + \frac{(1-\omega)^2 y^2}{8\beta^2\theta (G_r - G_l)^3} = 0.$$

Adding the two FOCs gives,

$$\frac{\omega}{2\sqrt{G_l^*}} = \frac{\omega}{2\sqrt{1-G_r^*}},$$

which implies that the equilibrium is unique and symmetric: $G_l^* = 1 - G_r^* = G^*$ where G^* is given by

$$\frac{1}{2} + \frac{\omega}{2\sqrt{G^*}} = \frac{(1-\omega)^2 y}{8\beta^2\theta (1-2G^*)^3}.$$

Since LHS of the above equation is strictly decreasing in G^*, and the RHS is strictly increasing in G^* for $G^* \leq 1/2$, therefore, there is at most one solution. □

A.4.2 Three-Regime Continent

For $S = 3$, we define the borders as

$$\begin{aligned} b_{1,2} &= \frac{G_1 + G_2}{2} + (1-\omega)\frac{M_1 - M_2}{2\beta (G_2 - G_1)} + \omega\left(\sqrt{G_1}\right), \text{ and} \\ b_{2,3} &= \frac{G_2 + G_3}{2} + (1-\omega)\frac{M_2 - M_3}{2\beta (G_3 - G_2)} - \omega\left(\sqrt{1-G_3}\right). \end{aligned}$$

The following lemma shows the unique equilibrium is symmetric.

Lemma 5. *In the absence of external threats, the equilibrium is unique and symmetric. Military expenses are the same for the two regimes closer to the frontier but higher for the regime in the middle such that*

$$\begin{aligned} M_1^* &= \frac{y(1-\omega)}{4\beta\theta (G_2 - G_1)}, \\ M_2^* &= \frac{y(1-\omega)(G_3 - G_1)}{4\beta\theta (G_3 - G_2)(G_2 - G_1)}, \quad \textit{and} \\ M_3^* &= \frac{y(1-\omega)}{4\beta\theta (G_3 - G_2)}, \end{aligned}$$

and the equilibrium locations of capitals are symmetric such that $G_1 = 1 - G_3 = G^*$, $G_2 = 1/2$ *where*

$$\frac{1}{2} + \frac{\omega}{2\sqrt{G^*}} - \frac{2y(1-\omega)^2}{\beta^2\theta (1-2G^*)^3} = 0.$$

Proof. First, in the second stage,

$$\max_{M_1} V_1 = b_{1,2} y - \theta M_1^2,$$
$$\max_{M_2} V_2 = (b_{2,3} - b_{1,2})\, y - \theta M_2^2, \text{ and}$$
$$\max_{M_3} V_3 = (1 - b_{2,3})\, y - \theta M_3^2.$$

The FOCs are

$$\frac{(1-\omega)}{2\beta (G_2 - G_1)} y - 2\theta M_1 = 0,$$
$$\frac{(1-\omega)\, y}{2\beta (G_3 - G_2)} + \frac{(1-\omega)\, y}{2\beta (G_2 - G_1)} - 2\theta M_2 = 0, \text{ and}$$
$$\frac{(1-\omega)}{2\beta (G_3 - G_2)} y - 2\theta M_3 = 0.$$

Hence,

$$M_1^* = \frac{y (1-\omega)}{4\beta\theta (G_2 - G_1)},$$
$$M_2^* = \frac{y (1-\omega)}{4\beta\theta (G_3 - G_2)} + \frac{y (1-\omega)}{4\beta\theta (G_2 - G_1)} = \frac{y (1-\omega)(G_3 - G_1)}{4\beta\theta (G_3 - G_2)(G_2 - G_1)}, \text{ and}$$
$$M_3^* = \frac{y (1-\omega)}{4\beta\theta (G_3 - G_2)}.$$

Note that

$$M_1^* - M_2^* = -\frac{y (1-\omega)}{4\beta\theta (G_3 - G_2)},$$
$$M_2^* - M_3^* = \frac{y (1-\omega)}{4\beta\theta (G_2 - G_1)}.$$

Now,

$$b_{1,2}^* = \frac{G_1 + G_2}{2} + \frac{-y (1-\omega)^2}{8\beta^2\theta (G_3 - G_2)(G_2 - G_1)} + \omega\sqrt{G_1}, \text{ and}$$
$$b_{2,3}^* = \frac{G_2 + G_3}{2} + \frac{y (1-\omega)^2}{8\beta^2\theta (G_3 - G_2)(G_2 - G_1)} - \omega \left(\sqrt{1 - G_3}\right).$$

Note that

$$b_{2,3}^{*}-b_{1,2}^{*}=\frac{G_3-G_1}{2}+\frac{y\left(1-\omega\right)^2}{4\beta^2\theta\left(G_3-G_2\right)\left(G_2-G_1\right)}-\omega\left(\sqrt{1-G_3}\right)-\omega\sqrt{G_1}.$$

Moreover,

$$\begin{aligned}
\frac{db_{1,2}}{dG_1}&=\frac{1}{2}+\frac{-y\left(1-\omega\right)^2}{8\beta^2\theta\left(G_3-G_2\right)\left(G_2-G_1\right)^2}+\frac{\omega}{2\sqrt{G_1}},\\
\frac{d\left(b_{2,3}-b_{1,2}\right)}{dG_2}&=\frac{-y\left(1-\omega\right)^2\left(\left(G_3-G_2\right)-\left(G_2-G_1\right)\right)}{8\beta^2\theta\left(G_3-G_2\right)^2\left(G_2-G_1\right)^2},\\
\frac{db_{2,3}}{dG_3}&=\frac{1}{2}-\frac{y\left(1-\omega\right)^2}{8\beta^2\theta\left(G_3-G_2\right)\left(G_2-G_1\right)}+\frac{\omega}{2\sqrt{1-G_3}},
\end{aligned}$$

and

$$\begin{aligned}
\frac{d\left(\theta M_1^2\right)}{dG_1}&=2\theta\left(\frac{y\left(1-\omega\right)}{4\beta\theta}\right)^2\frac{1}{\left(G_2-G_1\right)^3}=\frac{y^2\left(1-\omega\right)^2}{8\beta^2\theta\left(G_2-G_1\right)^3},\\
\frac{d\left(\theta M_2^2\right)}{dG_2}&=2\theta\left(\frac{y\left(1-\omega\right)\left(G_3-G_1\right)}{4\beta\theta}\right)^2\frac{-\left(\left(G_3-G_2\right)-\left(G_2-G_1\right)\right)}{\left(G_3-G_2\right)^3\left(G_2-G_1\right)^3},\\
\frac{d\left(\theta M_3^2\right)}{dG_3}&=2\theta\left(\frac{y\left(1-\omega\right)}{4\beta\theta}\right)^2\frac{-1}{\left(G_3-G_2\right)^3}=\frac{-y^2\left(1-\omega\right)^2}{8\beta^2\theta\left(G_3-G_2\right)^3}.
\end{aligned}$$

Now, in the first stage,

$$\begin{aligned}
&\max_{G_1}V_1=b_{1,2}^{*}y-\theta M_1^{2*},\\
&\max_{G_2}V_2=\left(b_{2,3}^{*}-b_{1,2}^{*}\right)y-\theta M_2^{2*},\ \text{and}\\
&\max_{G_3}V_3=\left(1-b_{2,3}^{*}\right)y-\theta M_3^{2*}.
\end{aligned}$$

The FOCs are

$$\begin{aligned}
\left(\frac{1}{2}+\frac{-y\left(1-\omega\right)^2}{8\beta^2\theta\left(G_3-G_2\right)\left(G_2-G_1\right)^2}+\frac{\omega}{2\sqrt{G_1}}\right)y-\frac{y^2\left(1-\omega\right)^2}{8\beta^2\theta\left(G_2-G_1\right)^3}&=0,\\
-\frac{y^2\left(1-\omega\right)^2\left(\left(G_3-G_2\right)-\left(G_2-G_1\right)\right)}{4\beta^2\theta\left(G_3-G_2\right)^2\left(G_2-G_1\right)^2}+\frac{y^2\left(1-\omega\right)^2\left(G_3-G_1\right)^2}{8\beta^2\theta}\frac{\left(G_3-G_2\right)-\left(G_2-G_1\right)}{\left(G_2-G_1\right)^3\left(G_3-G_2\right)^3}&=0,\ \text{and}\\
\left(\frac{1}{2}-\frac{y\left(1-\omega\right)^2}{8\beta^2\theta\left(G_3-G_2\right)\left(G_2-G_1\right)}+\frac{\omega}{2\sqrt{1-G_3}}\right)y+\frac{y^2\left(1-\omega\right)^2}{8\beta^2\theta\left(G_3-G_2\right)^3}&=0.
\end{aligned}$$

Adding up the the first and third FOCs gives us

$$\frac{\omega}{2\sqrt{G_1^*}} = \frac{\omega}{2\sqrt{1-G_3^*}}.$$

Hence,

$$G_1^* = 1 - G_{3.}^*$$

As for the second FOC,

$$\frac{y^2(1-\omega)^2\left((G_3^*-G_2^*)-(G_2^*-G_1^*)\right)\left(-2(G_2^*-G_1^*)(G_3^*-G_2^*)+(G_3^*-G_1^*)^2\right)}{8\beta^2\theta(G_3^*-G_2^*)^3(G_2^*-G_1^*)^3} = 0.$$

Then,

$$\left((G_3^*-G_2^*)-(G_2^*-G_1^*)\right)\left((G_2^*-G_1^*)^2+(G_3^*-G_2^*)^2\right) = 0.$$

Hence,

$$G_3^* - G_2^* = G_2^* - G_1^*.$$

Since $G_1^* = 1 - G_3^*$, $G_2^* = \frac{1}{2}$. Let $\Delta G = G_3^* - G_2^* = G_2^* - G_1^* = \frac{1-2G^*}{2}$. Then,

$$\left(\frac{1}{2} + \frac{-y(1-\omega)^2}{8\beta^2\theta(\Delta G)^3} + \frac{\omega}{2\sqrt{G^*}}\right)y - \frac{y^2(1-\omega)^2}{8\beta^2\theta(\Delta G)} = 0,$$

or

$$\frac{1}{2} + \frac{\omega}{2\sqrt{G^*}} = \frac{y(1-\omega)^2}{4\beta^2\theta\left(\frac{1-2G^*}{2}\right)^3}.$$

Since the LHS of the above equation is strictly decreasing in G^*, and the RHS is strictly increasing in G^* for $G^* \leq 1/2$, there is at most one solution. □

A.5 Multiple Military Bases

In Section 3, we assume that every regime can only set up one military base, which we referred to as the capital city. In this subsection we consider what happens if we allow a regime to set up multiple military bases.

It is clear that if there is no additional cost to set up new military bases, the empire or political unification will always dominate interstate competition in terms of stability in the face of threats. This is because the empire suffers no handicap in multitasking, since it is free to set up auxiliary military bases outside it capital city. On top of this, unlike interstate competition, the empire does not need to expend resources to compete with other regimes—since it has no competitors.

Historically, one important constraint that empires faced in setting up auxiliary armies is the risk of military usurpation. Once empowered, commanders of the auxiliary armies often sought to replace the incumbent rulers (See the discussion in Section 5.1).

Another constraint is the fixed cost of setting up new military bases. Note that if the fixed cost is large enough, this alone would justify our "one military per regime" assumption.

But suppose that the fixed cost of setting up a new military base is zero. And suppose that a regime may set up any number of auxiliary military bases outside its capital city. To model agency cost in the form of military usurpation, we assume that to prevent a military usurpation, the strength of the main army projected from the capital city (where the central government is located) has to be no less than the strength of the auxiliary army at the auxiliary military base.

Formally, preventing military usurpation requires $m\left(M_e, |G_e - G_a|\right) \geq m\left(M_a, 0\right)$, where G_a and M_a are the location and military investment of the auxiliary military base. We now show that it is optimal for an empire not to build auxiliary military bases.

Claim 5. *Regime e never maintains more than one military base.*

Proof. Under a one-sided threat, it is clear that optimality requires regime e to maintain only one military base.

Now consider a two-sided threat. Following Proposition 1, we consider three cases: (1) $\Lambda \leq \Lambda_I$; (2) $\Lambda_I < \Lambda \leq \Lambda_{II}$, and (3) $\Lambda > \Lambda_{III}$. Case (1): since regime e does not need to invest on the military, there is no need to set up auxiliary military bases. Case (2): since regime e only deals with the threat from one side, there is no need to set up another military base. Case (3): if regime e set up only one military base, it will be located at the center of continent to deal with the threat from both sides. If two military bases are preferred to one, the capital city and the auxiliary military base will be located to protect one frontier each. To minimize investment cost, the regime will invest just enough on the auxiliary army to block the threat at the location of the auxiliary military base. However, to satisfy the no military usurpation constraint, the military strength of the main army projected from the capital city has to be no less than the military strength of the auxiliary army at the location of the auxiliary military base. This implies that setting up the auxiliary military base is suboptimal. $\square$

The proof for regimes under interstate competition is similar.

A.6 Further Empirical Evidence

Here, we test the robustness of the results presented in Section 4 by replacing *Fragmentation* with *#Regime* (the *log* number of regimes in China Proper in decade t. See Table 1) as the dependent variable. The results obtained are consistent with those discussed in in Section 4.

Tables 7 and 8 are analogous to Tables 2 and 3 in Section 4. For example, column (a) of Table 7 suggests that while nomadic attacks appear to have negligible effect on the number of regimes in China proper, the lagged effect is significant: every additional nomadic attack is associated with a decrease in the number of regimes in China proper one decade later by 1.37%. In the VAR specification, the corresponding estimate is -1.30%.

Table 7: ADL Model

	(a)	(b)
Dependent variable:	#Regimes	#Regimes
#Regimes: Lag 1	1.061***	1.019***
	(0.0664)	(0.0678)
#Regimes: Lag 2	-0.337***	-0.325***
	(0.0901)	(0.0901)
#Regimes: Lag 3	0.146**	0.113*
	(0.0635)	(0.0646)
Nomadic attacks	0.00379	0.00409
	(0.00614)	(0.00642)
Nomadic attacks: Lag 1	-0.0137**	-0.0126**
	(0.00614)	(0.00634)
Additional controls	No	Yes
Observations	203	203
R-squared	0.781	0.798
AIC	0.158	0.184

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table 8: VAR Model

	(a)	(b)
Dependent variable:	#Regimes	Nomadic attacks
#Regimes: Lag 1	1.026***	1.904***
	(0.0660)	(0.721)
#Regimes: Lag 2	-0.343***	-1.330
	(0.0882)	(0.963)
#Regimes: Lag 3	0.131**	-0.254
	(0.0638)	(0.696)
Nomadic attacks: Lag 1	-0.0130**	0.333***
	(0.00626)	(0.0684)
Nomadic attacks: Lag 2	0.000943	0.236***
	(0.00654)	(0.0715)
Nomadic attacks: Lag 3	-0.00269	0.000803
	(0.00632)	(0.0690)
Additional controls	Yes	Yes
Observations	203	203

A.7 Population Fluctuations

Pop. ('000)	t	t^2	t^3	t^4	t^5	t^6	N	Adj. R^2
China	69.4***	-	-	-	-	-	14	0.58
Europe	54.2***	-	-	-	-	-	14	0.67
China	-77.8	0.081***	-	-	-	-	14	0.78
Europe	-68.8***	0.068***	-	-	-	-	14	0.94
China	44.2	-0.092	$6.4 \cdot 10^{-5}$	-	-	-	14	0.79
Europe	-0.62	-0.029	$3.6 \cdot 10^{-5}$*	-	-	-	14	0.95
China	-144.9	0.44	$-4.1 \cdot 10^{-4}$	$1.3 \cdot 10^{-7}$	-	-	14	0.80
Europe	-12.0	0.026	$7.1 \cdot 10^{-6}$	$8.1 \cdot 10^{-9}$	-	-	14	0.95
China	300.5	-1.61	0.0027	$-1.9 \cdot 10^{-6}$*	$4.6 \cdot 10^{-10}$*	-	14	0.86
Europe	144.7	-0.72	0.0011	$-7.1 \cdot 10^{-7}$	$1.6 \cdot 10^{-10}$*	-	14	0.96
China	242.3	-1.21	0.0018	$-8.9 \cdot 10^{-7}$	$-5.0 \cdot 10^{-11}$	$9.6 \cdot 10^{-14}$	14	0.84
Europe	34.4	0.036	$-6.7 \cdot 10^{-4}$	$1.2 \cdot 10^{-6}$	$-8.1 \cdot 10^{-10}$	$1.8 \cdot 10^{-13}$	14	0.96

Constant terms are not reported. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 9: Fitting Year Polynomials to Chinese and European Population Data. Adjusted R^2 is higher for Europe than for China in each case. See discussion in Section 5.3.