JOURNAL OF REGIONAL SCIENCE, VOL. 49, NO. 2, 2009, pp. 243–261

EXPLORING THE NUMBER OF FIRST-ORDER POLITICAL SUBDIVISIONS ACROSS COUNTRIES: SOME STYLIZED FACTS*

Maximilian Auffhammer

Department of Agricultural and Resource Economics, 207 Giannini Hall, University of California, Berkeley, CA 94720-3310. E-mail: auffhammer@berkeley.edu

Richard T. Carson

Department of Economics, University of California, San Diego, 9500 Gilman Drive #0508, La Jolla, CA 92093-0508. E-mail: rcarson@weber.ucsd.edu

ABSTRACT. Why do some countries have no first-order administrative subdivisions (e.g., states or provinces), whereas other countries have over 80? Recently, economists have started to look at the optimal size of countries and forces influencing the creation of local political jurisdictions like school districts. This paper provides the first analysis of the "missing middle" level of political jurisdictions common to all countries. We empirically examine how country size, natural transportation infrastructure, location, population fractionalization, and level of development affect the number of first-order subdivisions. The number of first-order subdivisions is shown to be associated in a nonlinear way with measures of fractionalization—exhibiting a U-shaped Kuznets curve for ethnic heterogeneity and an inverted Kuznets curve for lingual and religious heterogeneity. This is a different and more complex relationship than that found for local political jurisdictions where greater heterogeneity is associated with more districts suggesting that first-order political subdivisions may serve a different role.

1. INTRODUCTION

Many nation states have historically split their land area into smaller subdivisions (e.g., shires, states, cantons). Central governments delegate the provision of certain public goods as well as the design and implementation of some policies to these lower levels of government at varying degrees. In the more distant past, these "state" governments were better able to collect taxes and supervise agricultural production. More recently, they have been charged with providing a range of government services such as education, health care, and local environmental quality. By the turn of the millennium, 197 countries had a total of 2,950 first-order subdivisions. The number of these first-order subdivisions across countries displays great variability, ranging from no

*We thank Matthew Kahn and two anonymous referees for constructive comments. Roger Gordon, Barry Naughton, Valery Ramey, James Rauch, Elisabeth Sadoulet, and Jeffrey Vincent provided helpful comments on an earlier draft. Any errors are ours.

Received: May 2007; revised: March 2008; accepted: May 2008.

DOI: 10.1111/j.1467-9787.2008.00583.x

such divisions to 89. The goal of this paper is to empirically examine which characteristics of countries explain the observed variation in the number of first-order political subdivisions across countries.

We use predictions from the rich theoretical literature on the number and size of countries and "states" as a motivation for our empirical model (Alesina and Spolaore, 1997, 2003). This literature predicts that the number of divisions is increasing in the degree of population heterogeneity and decreasing in the cost of provision of local public goods. Alesina, Hoxby, and Baqir, (2004) confirm these theoretical predictions by examining the number of school districts across U.S. counties. In this paper, we utilize a newly assembled data set on the number of first-order subdivisions for 197 countries. We make two specific contributions to the empirical literature on political subdivisions.

First, using a rich set of variables available across countries, we examine the relative impact of country size, absolute and relative location, natural transportation infrastructure, population heterogeneity, and level of development on the number of first-order political subdivisions. The estimated partial correlations are intended to serve as a set of stylized facts, rather than evidence of causal impacts.

Second, we more broadly examine the effect of population heterogeneity on first-order subdivisions. By examining the assumed to be linear impact of lingual, religious, ethnic, and income heterogeneity separately, we find that consistent with the literature, ethnic heterogeneity leads to a larger number of subdivisions. We do, however, find a statistically significant and negative impact of religious, lingual, and income heterogeneity. Further, when relaxing the assumption that the number of first-order subdivisions is linear in these indicators of heterogeneity, we find strong evidence of a nonlinear relationship between ethnic, lingual, and religious fragmentation and the number of first-order subdivisions. This nonlinear relationship resembles a Kuznets curve for ethnic fragmentation and an inverted Kuznets curve for lingual and religious fragmentation. The estimated turning point for all three measures occurs within sample.

The remainder of the paper is organized as follows. The next section surveys the relevant literature and motivates our reduced form empirical model. Section 3 provides the empirical model. Section 4 shows and discusses results from the count data estimation. Section 5 concludes.

2. BACKGROUND AND LITERATURE

While considering first-order subdivisions as endogenous in the short run may be difficult, there is anecdotal evidence that in the long run countries reorganize their internal structure. One example of increasing the number of subdivisions, while holding the land area fixed, is the Peoples' Republic of China in 1997 splitting Sichuan into two political subdivisions (Sichuan and Chongquing). At the time of the split, Sichuan was the most populous province in China and experiencing rapid population growth. Another example is Turkey, which, since 1957, has sequentially increased the number of provinces by 24. Nigeria, New Zealand, Uganda, Algeria, and most of the states of the former Soviet Union have also reorganized their internal structure in recent history. Many countries observed today are simply an agglomeration of previous countries or shires. Examples of this are Germany, Spain and, up until the early 1990s, the Soviet Union. If these "historical subdivisions" were optimal, we would observe them as first-order subdivisions today. Some countries that arose from such agglomerations have decided to reassign their first-order subdivisions by combining these historical shires or breaking them into smaller units. One example is Burgundy in France, which is a much smaller region than that ruled over by the count of Burgundy.[1] In France, there had been an extensive debate surrounding the internal reorganization into régions, since the *departements* were thought to be too small. Law (1999) provides an extensive overview with some historical detail on the internal structure of all nations.

Alesina and Spolaore (1997) provide a theoretical framework, which has spawned a sizeable literature on the optimal number of countries and subdivisions. Alesina and Spolaore (2003) provide an extensive overview and extension of this literature. They model a discrete set of types of individuals, who have preferences over the consumption of a number of public goods. Utility is increasing in the consumption of and decreasing in the distance from the public good. This distance could be interpreted as a measure of distance in tastes or actual physical distance. The key aspect of these models is that two individuals do not have to receive the same public good from a shared jurisdiction. Public goods are financed through individuals' tax payments to the budget constrained jurisdictions providing each public good. The local public good has a fixed cost component as well as a variable cost component, which is increasing in the number of individuals the public good is provided to. The social planner's choice variable is the number of subdivisions providing each public good. The equilibrium conditions in these models usually imply that the optimal number of subdivisions is increasing in the number of types of individuals as well as the distance of the average individual from the public good, which is often referred to as the heterogeneity cost. The number of subdivisions is decreasing in the fixed cost of setting up a jurisdiction to provide each public good. Countries and first-order subdivisions arise, if one adds a "hierarchy constraint." Here two individuals getting the first good from the same jurisdiction have to get the second good from the same jurisdiction, but not vice versa. The equilibrium prediction is therefore that larger countries with a more heterogenous population and a lower cost of public goods provision will have a larger number

[1]Burgundy was split into the duchy of Burgundy and the county of Burgundy. The duchy became the French province of Burgundy, whereas the county of Burgundy became Franche-Comté.

of subdivisions. While we do not formally test the predictions from this structural model, they provide some guidance as to which variables to include in our empirical application.

The relevant empirical literature at the highest level of aggregation explains the observed number of countries (e.g., Alesina and Spolaore, 2003). In the public finance literature, Alesina et al. (2004) apply the Alesina and Spolaore (1997) model to show a significant trade-off between economies of scale of public goods provision and racial heterogeneity in the determination of the number of school districts across U.S. counties. Alesina, Baqir, and Easterly (1999) provide empirical evidence of a similar trade-off when looking across cities.

In addition to this political economy literature, urban and regional economists have long studied the emergence and development of regional and urban economies. While we could not find evidence of any papers studying the number of first-order subdivisions in this literature, the number and size distribution of cities has been widely studied (e.g., Krugman, 1993; Gabaix and Ioannides, 2004; Kim and Margo, 2004). The empirical literature is focused on explaining the distribution of city size with a focus on the relative importance of increasing returns, congestion, trade, and nonmarket forces (Gabaix and Ioannides, 2004).

In the field of industrial organization, there is a literature dealing with the optimal number and size of firms in a market based on classic papers by Coase (1937) and Lucas (1978). The recent focus in this literature has been on deviations from the optimal number of firms in a market power context (Berry and Waldfogel, 1999). At a lower level of aggregation, there are a few papers looking at the number of plants for a multiplant firm (e.g., Chambers, 1998). In summary, these papers address the number of subdivisions at the largest (countries and firms) and the smallest (school districts, cities, and plants) level of aggregation. The current paper provides the first empirical study of the "missing middle," namely the optimal number of first-order political subdivisions across countries. The firm equivalent level of aggregation would be dividing firms into divisions, supervising production at groups of plants.

3. DATA AND EMPIRICAL MODEL

We use the number of administrative regions as defined in the Central Intelligence Agency's World Fact Book for the year 2000 as our measure of the number of political subdivisions. We exclude municipalities and parishes since they do not really reflect the notion of a "political subdivision" in the sense of this paper. We cross-checked the CIA definitions with the authoritative book on the subject "Administrative Subdivisions of Countries" by Law (1999). In the case of conflicting definitions, we consulted the national statistical agencies which, without exception, sided with Law (1999). In these cases, we therefore used Law's (1999) definition. The number of first-order political subdivisions

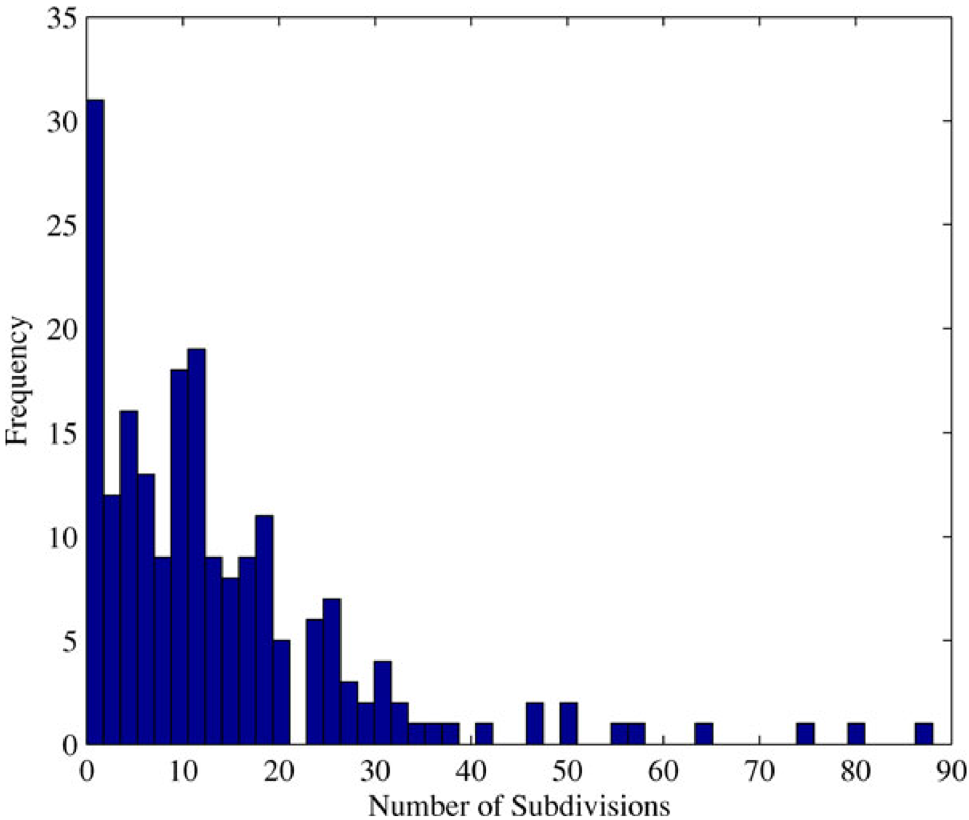

FIGURE 1: Empirical Distribution of First-Order Subdivisions.

is quite variable across countries and ranges from 1 to 89.[2] The distribution of the number of political subdivisions across countries is given in Figure 1.[3]

In order to establish the relative impact of country size, natural transportation infrastructure, location, population fractionalization, and level of development on the observed number of subdivisions, it would be optimal to have spatially and temporally randomly assigned treatments of these factors across countries. Lacking such an experiment and constrained by data limitations, we resort to exploring the cross-section of countries.[4] While drawing causal

[2]A country with one subdivision would indicate that the country has no sub-national political structure or administrative regions.

[3]The raw data, which include the type of subdivision we considered (e.g., state, district, oblast), are available upon request from the authors.

[4]We have explored the option of using a panel data set, but although there are changes in the number of subdivisions for quite a few countries, there is not enough temporal variation allowing us to identify the effects we are after using within country variation.

inference from our analysis is not recommended, by careful econometric reduced form modeling, we provide estimates of the relative importance of these factors in terms of partial correlations. More importantly, we will show that a more flexible functional form of the estimating equation results in an improved fit and leads to new and interesting conclusions as to how heterogeneity relates to the number of subdivisions.

The model in Alesina and Spolaore (2003) suggests that the number of divisions is uniformly increasing in the population size of a country. We include the area and population of each country as explanatory variables. Area proxies for physical distance between a central government and individuals. Since we control for population separately, it controls for the density of population for a given country. We would expect a positive and significant coefficient estimate on both variables. In addition, one could expect that for smaller countries it is not optimal to incur the fixed costs of setting up two separate layers of government. For smaller countries, it may therefore be optimal to provide all public goods from a unified central government. In the empirical analysis, we will allow for such a threshold effect by allowing the "no divisions" outcome to be drawn from two separate regimes.

As Kim and Margo (2004) point out, accessibility of the hinterland has been thought by some scholars to play a crucial role in the ability to settle the interior of the country. Good natural transportation infrastructure provided ways not only to settle the hinterland, but also to ship agricultural commodities and goods to transportation hubs, such as harbors and railway centers. Harbors specifically attract dense settlement which in turn requires the formation of multiple divisions. From a military, strategic point of view, it may be optimal to have the harbor controlled by a unified command, rather than split among several. Further, harbors and the rivers that feed them are natural borders, delineating one division from another. There are numerous examples of harbors which are located near the borders of political subdivisions. The large natural harbors, New York and Chesapeake Bay, are all located near state borders. In Europe, Le Havre is located at the border between Haute-Normandie and Basse-Normandie. Rotterdam harbor is located at the junction of three political subdivisions. We would therefore *a priori* expect countries with a higher number of natural harbors to have a larger number of subdivisions. To measure internal natural infrastructure, we also calculate the length of waterways in a country. Both measures could be interpreted as proxies for lower cost of local public goods provision, which Alesina and Spolaore (2003) predict increases the number of subdivisions.

Next, we examine the role of relative and absolute location of countries on the number of subdivisions. While there are numerous papers examining the role of, e.g., openness to trade (Alesina and Wacziarg, 1998) or international conflict (Alesina and Spolaore, forthcoming) on the number of countries, there is little theoretical guidance on how these factors affect divisions at the subnational level. In this paper, we model the choice of subdivisions conditional on country size.

Relative location in this paper refers to where a country is located in relation to other countries. Two simple measures of relative location are the length and number of borders with its neighbors. Historically, countries like Austria, which borders seven countries, would be expected to have a larger number of subdivisions along the border, in order to establish ownership of the land and its resources at the border. The intuition here is that many small "posts" along a border also allow one to defend more easily against invasions. We also control for whether a country is landlocked or not. One would expect a landlocked country to have a larger number of subdivisions, since it will need to defend its people and resources from surrounding neighbors vying for its territory. The effect would be the opposite for islands, or countries with lengthy coastlines, since there are no directly adjacent neighbors who would spill over into the nation's territory and take ownership of the resources in the border territories. The effect of border length, conditional on the number of neighbors, could go in either direction. Longer borders are more easily guarded by a larger number of subdivisions. If there are fixed costs to establishing institutions to guard the local borders, one could expect countries with longer borders having fewer subdivisions. As our measure of absolute location we include a dummy for which continent a country is located on, indicating the number of subdivisions relative to being located in Europe.

As our measures of country level population heterogeneity, we include the indicators of lingual, ethnic, and religious diversity of a country as provided by Alesina et al. (2003). These indexes of heterogeneity are calculated by subtracting the sum of squared shares of a certain group in the total population from one. A larger value for any of these indicators indicates a larger degree of fractionalization, which is equivalent to a more heterogeneous population. According to the Alesina and Spolaore (2003) model, the number of subdivisions should linearly increase with increasing heterogeneity. The intuition behind this argument is that like individuals have similar preferences for certain public goods. Further, they may not wish to contribute tax payments to jurisdictions which provide services to individuals of a different type. An interesting question is at what level of aggregation this effect takes place. Members of different ethnic (lingual, religious) groups tend to live in the same neighborhoods within cities. In the United States, for example, we do not experience an easily observable separation of ethnicity (language, religion) by state, yet we see a clear separation by neighborhoods within urban areas. One only needs to stroll through the different neighborhoods of New York City or Los Angeles and experience this sorting. If this sorting only takes place at a city level of aggregation (Alesina et al., 1999), we may not be able to detect the effect or estimate a very small effect. As a final measure of heterogeneity not provided in Alesina et al. (2003), we include a measure of income inequality via a Gini coefficient for a small set of countries (World Bank, 2000).

Finally, consistent with Alesina et al. (2004), we use average income. In their paper, they use it as a proxy for how much individuals like the public good. Here we interpret this variable as an indicator of economic development.

TABLE 1: Summary Statistics (Year 2000)

Variable	Obs.	Mean	Std. Dev.	Min	Max
Regions	197	13.93	14.62	0	88.00
Population (10 million)	197	3.18	12.25	0.00	129.00
Area (100,000 km^2)	197	6.85	19.08	0.00	171.00
Africa	197	0.26	0.44	0	1
Asia	197	0.23	0.42	0	1
Europe	197	0.22	0.42	0	1
North America	197	0.14	0.35	0	1
South America	197	0.06	0.24	0	1
Oceania	197	0.08	0.27	0	1
Land border length (1,000 km)	197	2.51	3.31	0	22.15
Number of land borders	197	3.14	2.79	0	17.00
Landlocked	197	0.20	0.40	0	1.00
Coastline (1,000 km)	197	3.84	15.93	0	202.08
Harbors	197	5.75	5.26	0	32.00
Waterways (1,000 km)	197	3.22	11.65	0	110.00
Island	197	0.23	0.42	0	1.00
Ethnic	185	0.44	0.26	0	0.93
Language	180	0.39	0.28	0	0.92
Religion	190	0.44	0.23	0	0.86
GDP p.c. (1,000 PPP US$)	195	8.72	9.54	0.55	44.00
Gini	110	39.48	9.75	21.70	62.90

One could potentially expect that more developed countries with better quality institutions are more efficient at providing public goods and therefore require a smaller number of first-order subdivisions. We provide summary statistics of all variables used in estimation in Table 1.

We define our dependent variable $Regions_i$ as the number of observed first-order subdivisions minus one. A country with no first-order subdivisions would therefore be observed as a zero in the data. Since the dependent variable is a discrete count variable, bounded by zero, the appropriate estimation techniques are maximum likelihood-based Poisson estimation. In this case, the assumed underlying data-generating process assumes that each count is generated by the same Poisson process.

Sixteen percent of the countries in our sample have no subdivisions. In our sample, 88 percent of the zero observations come from countries that have a land area smaller than that of Massachusetts. Seventy-six percent of the zero observations have a population less than one million people. A zero inflated Poisson (ZIP) model (Lambert, 1994) will allow us to control for the possibility that the zero observations may be drawn from two regimes. The first regime producing zeros is a size regime, where countries will only have subdivisions if they are larger than a certain threshold, as argued above. The second regime producing zero observations is the regular Poisson process. We let

$Regions_i = 0$ with probability ψ_i and let $Regions_i$ be distributed as Poisson λ_i with probability $(1 - \psi_i)$. We define the Zero Inflated Poisson model as follows:

$$P[Regions_i = 0] = \psi_i + (1 - \psi_i)R_i(0), \tag{1}$$

$$P[Regions_i = j] = (1 - \psi_i)R_i(j), \tag{2}$$

where $R_i(Regions = j) = \frac{e^{-\lambda_i} \cdot \lambda_i^j}{j!}$ and $\lambda_i = e^{\beta' x_i}$. We model the state probability to be distributed as $\psi_i \sim \text{logistic}(\theta\ z_i)$. The z_i here are variables, which determine the probability of a zero being drawn from the first regime. The link function for the first regime is specified as a logit and the full model is estimated by maximum likelihood using data on 197 countries. As a robustness check we estimate both the regular Poisson model and the ZIP model. We then conduct a Vuong test to decide which model more accurately approximates the true data-generating process.

4. ESTIMATION RESULTS

In this section, we present the results from the regular and zero inflated Poisson estimation on the cross-section of 197 countries. Columns (1)–(3) of Table 2 show the estimation results from the regular Poisson estimation. Model (1) includes the full set of covariates, yet the smallest number of observations, due to the limited number of observations on income equality. Model (2) removes income and income inequality from the estimation and model (3) removes the fractionalization variables.

The variables capturing country size carry signs consistent with prior expectations. The parameter estimate on population is positive and significant in all three models, suggesting an increased number of subdivisions for countries with larger populations. The coefficient on land area is positive and significant in model (1), yet not significant in models (2) and (3). This may be due to the high correlation between the measure of land area and population. The point estimates are stable and positive in all models.

The natural transportation infrastructure results are partially consistent with our expectations and the literature. The number of harbors is significant at the 1 percent level in all models. At the sample average, each harbor is consistent with approximately one additional subdivision. The length of waterways carries a surprising negative sign. This measure differs from the number of streams variable used by Alesina et al. (2004). Our measure treats a country with a single 1,000 km long waterway the same as a country with ten 100 km long waterways. We would expect the latter country to have a larger number of subdivisions relative to the former.

The parameter estimates on the relative location variables have the expected signs with the exception of border length. The island and landlocked dummies carry opposite signs, suggesting a larger number of subdivisions for landlocked countries and a smaller number for islands, which is consistent with

TABLE 2: Estimates from the Poisson (1-3) and Zero Inflated Poisson (4-6) (ZIP) Model

	(1)	(2)	(3)	(4)†	(5)	(6)
Population	0.010	0.005	0.005	0.009 [0.17]	0.005	0.004
	(0.002)***	(0.002)**	(0.002)**	(0.002)***	(0.002)**	(0.002)*
Area	0.011	−0.0007	0.0008	0.010 [0.18]	0.002	0.003
	(0.003)***	(0.003)	(0.003)	(0.003)***	(0.003)	(0.002)
Harbors	0.062	0.069	0.069	0.062 [1.14]	0.054	0.051
	(0.007)***	(0.005)***	(0.005)***	(0.006)***	(0.006)***	(0.005)***
Waterways	−0.008	−0.008	−0.011	−0.007 [−0.13]	−0.005	−0.008
	(0.004)**	(0.003)**	(0.003)***	(0.004)*	(0.003)	(0.003)**
Landlocked	0.270	0.346	0.236	0.266 [5.24]	0.318	0.215
	(0.079)***	(0.068)***	(0.063)***	(0.079)***	(0.068)***	(0.063)***
Coast	−0.004	0.0003	−0.0004	−0.004 [−0.08]	−0.001	−0.002
	(0.002)**	(0.001)	(0.001)	(0.002)**	(0.001)	(0.001)
Island	−0.632	−0.620	−0.854	−0.477 [−7.24]	−0.039	−0.220
	(0.135)***	(0.102)***	(0.092)***	(0.138)***	(0.105)	(0.094)**
Border length	−0.076	0.002	0.0002	−0.070 [−1.28]	−0.008	−0.0002
	(0.020)***	(0.016)	(0.015)	(0.020)***	(0.015)	(0.015)
Number of borders	0.033	0.060	0.060	0.033 [0.24]	0.055	0.053
	(0.013)**	(0.012)***	(0.011)***	(0.013)**	(0.012)***	(0.011)***
Africa	0.015	0.189	0.223	0.006 [0.11]	−0.023	0.026
	(0.098)	(0.071)***	(0.063)***	(0.097)	(0.073)	(0.063)
North America	−0.061	0.259	0.166	0.034 [0.63]	0.266	0.205
	(0.135)	(0.096)***	(0.087)*	(0.137)	(0.095)***	(0.086)**
South America	−0.056	0.132	0.280	−0.045 [−0.82]	−0.016	0.130
	(0.141)	(0.100)	(0.093)***	(0.141)	(0.100)	(0.092)
Asia	0.169	0.191	0.299	0.148 [2.82]	0.047	0.163
	(0.091)*	(0.065)***	(0.061)***	(0.092)	(0.066)	(0.061)***

Continued

TABLE 2: Continued

	(1)	(2)	(3)	(4)†	(5)	(6)
Oceania	0.023	0.407	0.375	0.002 [.04]	0.025	−0.024
	(0.245)	(0.146)***	(0.142)***	(0.246)	(0.149)	(0.143)
Ethnic	0.610	0.141		0.599 [11.01]	0.315	
	(0.172)***	(0.134)		(0.170)***	(0.135)**	
Language	−0.244	−0.160		−0.232 [−4.27]	−0.145	
	(0.152)	(0.121)		(0.150)	(0.117)	
Religion	−0.813	−0.602		−0.799 [−14.69]	−0.695	
	(0.122)***	(0.096)***		(0.122)***	(0.097)***	
GDP	−0.011			−0.013 [−0.23]		
	(0.004)***			(0.004)***		
Gini	−0.004			−0.006 [−0.12]		
	(0.004)			(0.004)		
Constant	2.886	2.101	1.835	2.957	2.340	2.127
	(0.172)***	(0.090)***	(0.078)***	(0.174)***	(0.091)***	(0.080)***
Population				−0.146	0.568	0.569
				(4.822)	(2.312)	(2.310)
Area				−12.987	−16.544	−16.547
				(13.115)	(5.704)***	(5.704)***
Constant				0.236	0.932	0.932
				(1.885)	(0.460)**	(0.460)**
Obs.	106	175	197	106	175	197
ln(L)	−635.13	−1,044.94	−1,176.32	−627.43	−911.96	−1,002.72
No Subdivisions	1	19	30	1	19	30
Vuong Statistic				0.84	3.40***	3.56***

†Marginal Effects at the sample means are included in square brackets.
Standard Errors are reported in round brackets.*Significant at 10 percent; **significant at 5 percent; ***significant at 1 percent.

our prior expectations. Coast is measured in kilometers of coastline, which has a negative sign across all models and is significant only in the least parsimonious specification. The estimation results suggest that countries with more neighbors have a larger number of subdivisions. The length of the borders, however, results in a negative and sizeable effect on the number of subdivisions. This is surprising, yet one possible explanation is that there are increasing returns to guarding a longer border, *ceteris paribus*, and therefore this effect may result in larger and therefore fewer subdivisions. The parameter estimates on the continent dummy variables in model (1) are insignificant. However, the point estimates are positive and significant for models (2) and (3), which include a larger number of countries, suggesting a larger number of subdivisions for countries outside Europe.

The models controlling for population heterogeneity provide some interesting results. Ethnic heterogeneity has a positive and significant effect in all models. Religious heterogeneity is statistically significant and negative in all models. Language heterogeneity also has a negative effect, yet is not statistically significant in any model.[5] Overall these estimation results suggest that heterogeneity does matter, yet the type of heterogeneity one considers may lead to different predictions. If one considers religious heterogeneity, estimation results suggest the opposite of what is implied by Alesina and Spolaore (2003), namely that increased heterogeneity decreases the number of subdivisions in a country, *ceteris paribus*. The finding on religious heterogeneity is opposite of what Alesina et al. (2004) find in their study looking at school districts. Ethnic heterogeneity has the expected impact—an increase in the number of predicted subdivisions.

The last set of variables included account for per capita income and income inequality. GDP has a small, yet statistically significant, effect on the number of subdivisions of a country, which is consistent with Alesina et al. (2004) findings for school attendance areas. This finding is consistent with wealthier countries being more efficient at administering local public goods. Economic heterogeneity, as measured by the Gini coefficient, is not statistically different from zero and the estimated coefficient is very small and negative. We only have data on the Gini coefficient for little more than half of the sample and it is generally considered to be a badly measured variable, which is why we drop it from the remainder of the analysis.

Models (4)–(6) present the zero inflated Poisson estimation results. The bottom rows of Table 2 for these models show the results from the logit estimation of whether a country has any subdivisions. A success in this estimation is having no subdivisions, and we would therefore expect the coefficients on population and area to be negative. Population does not enter any of the three estimated models statistically different from zero in the first regime estimated

[5]The correlation coefficients between ethnic and lingual, ethnic and religious, and lingual and religious fractionalization are 0.71, 0.15, and 0.25 respectively.

via the logit link. The high correlation ($\rho_r = 0.79$) between the level population and land area may cause inflation of the standard errors.[6] The logit estimation results suggest that country area is the main determinant of whether a country has any subdivisions at all, since the estimated parameter is negative and statistically and economically significant in each but the first model. In order to test whether the ZIP model outperforms a traditional Poisson model, we conduct a Vuong test, which rejects the null of a simple Poisson process for all but the first model. The first model includes the Gini coefficient, which is not available for the smaller countries. The available sample consists countries all but one of which have first-order subdivisions. It is therefore not surprising that the Vuong test fails to reject the null of a simple Poisson for this model. Overall, the estimation results from the ZIP model are very similar to the regular Poisson process. The coefficients on island become smaller and the continent dummies become insignificant with the exception for North America's in models (5) and (6) and Asia in model (6). The estimates on the fractionalization variables are almost identical. Since the ZIP process fits the data slightly better, we will use it for the remainder of the paper.

Table 2 presents the estimation results of entering the z_i and the x_i linearly. It is reasonable to question whether the number of subdivisions may nonlinearly depend on population size. We conducted a likelihood ratio test of a model where population and area enter linearly, versus a model where they enter as natural logs. The hypothesis test rejects the null hypothesis of a log relationship at the 1 percent level. We further test the logarithmic specification by including powers between 0 and 1, similar to a Box-Cox test for nonlinearities. The search over nonlinearity parameters using the log likelihood as a selection criterion, suggests that population and area enter linearly. We therefore include these variables as levels.

We conducted a range of further robustness checks. We dropped the five largest outliers and re-estimated the model. None of the results change significantly. We further estimated all models by using a zero inflated negative binomial estimation technique, without any qualitative changes to the results. Ordinary Least Squares Estimation on the set of countries with first-order political subdivisions resulted in qualitatively similar results.

The opposing signs on the measures of population heterogeneity call for further inquiry into how these measures may affect the observed number of first-order subdivisions across countries. We therefore rerun the model with all variables, but exclude the income variables due to the limited size of the resulting sample. We include each heterogeneity measure linearly and then again as a second-degree polynomial allowing for nonlinearities. Finally, we include all three measures with their squared terms jointly. These estimation results are presented in Table 3 below. The estimation results show even stronger evidence of the opposite effect for ethnic versus religious and lingual heterogeneity.

[6]We alternatively estimated the state probability ψ_i via probit estimation. The results for all models are almost identical.

TABLE 3: ZIP Parameter Estimates: Exploring the Role Heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Population	0.003	0.005	0.004	0.005	0.005	0.005	0.003
	(0.002)	(0.002)**	(0.002)**	(0.002)**	(0.002)**	(0.002)**	(0.002)
Area	0.002	0.004	0.002	0.004	0.005	0.003	0.002
	(0.003)	(0.003)*	(0.003)	(0.003)*	(0.003)*	(0.003)	(0.003)
Landlocked	0.388	0.242	0.247	0.272	0.273	0.308	0.337
	(0.069)***	(0.065)***	(0.065)***	(0.066)***	(0.066)***	(0.066)***	(0.067)***
Coast	−0.000	−0.002	−0.001	−0.002	−0.002	−0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Island	−0.011	−0.042	−0.039	−0.255	−0.258	−0.167	−0.117
	(0.107)	(0.100)	(0.099)	(0.101)**	(0.101)**	(0.097)*	(0.098)
Harbors	0.058	0.055	0.057	0.059	0.058	0.058	0.056
	(0.006)***	(0.006)***	(0.006)***	(0.006)***	(0.006)***	(0.006)***	(0.006)***
Waterways	−0.002	−0.010	−0.008	−0.012	−0.012	−0.007	−0.005
	(0.004)	(0.003)***	(0.003)**	(0.003)***	(0.003)***	(0.003)**	(0.003)
Border length	−0.016	−0.005	−0.003	−0.006	−0.006	−0.005	−0.004
	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Number of borders	0.045	0.049	0.046	0.052	0.053	0.046	0.049
	(0.012)***	(0.012)***	(0.012)***	(0.012)***	(0.012)***	(0.012)***	(0.012)***
Africa	−0.288	−0.094	−0.001	−0.101	−0.111	−0.133	−0.295
	(0.096)***	(0.080)	(0.083)	(0.079)	(0.082)	(0.076)*	(0.083)***
North America	0.081	0.158	0.138	0.165	0.166	0.090	0.081
	(0.101)	(0.094)*	(0.094)	(0.095)*	(0.095)*	(0.092)	(0.092)
South America	−0.230	0.036	0.032	−0.037	−0.039	−0.083	−0.160
	(0.107)**	(0.099)	(0.099)	(0.099)	(0.099)	(0.099)	(0.100)
Asia	−0.136	0.041	0.045	0.031	0.032	−0.062	−0.126
	(0.077)*	(0.072)	(0.071)	(0.072)	(0.072)	(0.073)	(0.074)*
Oceania	−0.264	−0.292	−0.251	−0.109	−0.107	−0.035	−0.208
	(0.160)*	(0.151)*	(0.150)*	(0.153)	(0.153)	(0.151)	(0.155)

Continued

TABLE 3: Continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
GDP	−0.0130	−0.010	−0.009	−0.014	−0.014	−0.011	−0.014
	(0.003)***	(0.003)***	(0.003)***	(0.003)***	(0.003)***	(0.003)***	(0.003)***
Ethnic	1.709	−0.022	1.123				
	(0.467)***	(0.103)	(0.357)***				
Ethnic^2	−1.746		−1.341				
	(0.520)***		(0.400)***				
Language	−1.090			−0.222	−0.354		
	(0.363)***			(0.089)**	(0.286)		
Language^2	1.238				0.163		
	(0.421)***				(0.336)		
Religion	−2.445					−0.662	−2.479
	(0.391)***					(0.090)***	(0.359)***
Religion^2	2.197						2.277
	(0.464)***						(0.435)***
Constant	2.765	2.283	2.081	2.380	2.398	2.578	2.901
	(0.145)***	(0.100)***	(0.118)***	(0.096)***	(0.103)***	(0.100)***	(0.117)***
Population	0.569	0.201	0.201	1.089	1.089	0.971	0.971
	(2.311)	(2.666)	(2.666)	(1.621)	(1.621)	(2.065)	(2.065)
Area	−16.544	−17.826	−17.826	−16.366	−16.366	−18.186	−18.187
	(5.704)***	(6.063)***	(6.063)***	(5.451)***	(5.451)***	(5.943)***	(5.943)***
Constant	0.932	1.083	1.083	0.851	0.851	1.043	1.043
	(0.460)**	(0.435)**	(0.435)**	(0.436)*	(0.436)*	(0.411)**	(0.411)**
Obs.	175	185	185	179	179	189	189
ln(L)	−888.014	−969.89	−964.222	−950.49	−950.373	−949.486	−935.96
No subdivisions	19	23	23	20	20	25	25
Vuong statistic	3.40***	3.68***	3.73***	3.33***	3.33***	3.56***	3.60***

Standard errors are reported in round brackets. *Significant at 10 percent; **significant at 5 percent; ***significant at 1 percent.

In model (2) ethnic heterogeneity is included solely and enters linearly. It has a small negative, statistically insignificant effect. When we add the squared term, a highly significant “inverse-U” functional form emerges. This is similar to an ethnic Kuznets curve, whereby the number of divisions is increasing at low levels of heterogeneity and decreasing at high levels of heterogeneity. When we include religious heterogeneity, it is significant and negative when entered linearly, but once we include the higher order term a statistically significant “U-shape” emerges, whereby the number of subdivisions decreases at low levels of heterogeneity and then increases at higher levels. The same is true for lingual heterogeneity, yet the model rejects a nonlinear specification when this variable is included as a sole measure of fractionalization. When we include polynomials for all three measures jointly, the coefficients on the first- and second-order terms are all significant at the 1 percent level. The parameter estimates on lingual and religious heterogeneity are almost identical in the full model, suggesting that the omitted and highly correlated measure of ethnic heterogeneity in Equations (4) and (5) leads to a bias toward zero on the lingual heterogeneity coefficient. The estimated turning points for ethnic, lingual, and religious heterogeneity are at 0.489, 0.440, and 0.556, respectively, which is roughly at the midpoint of the sample for each variable. Figure 2 shows a scatter plot of ethnic versus religious heterogeneity labeling the observations by country. The lines indicate the estimated turning points for each of the two variables.

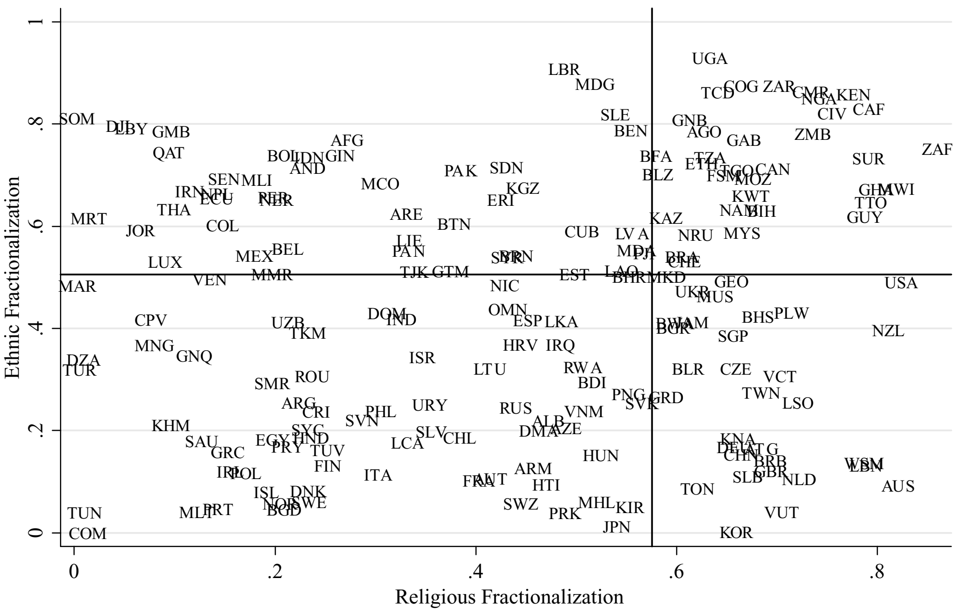

FIGURE 2: Distribution of Ethnic and Religious Fractionalization.

The countries that are experiencing a pull from the ethnic and religious heterogeneity in a common direction are on the diagonal (North-West and South-East Quadrant). High ethnic heterogeneity and low religious heterogeneity are consistent with a lower number of subdivisions. Low ethnic heterogeneity and high religious fragmentation combined with high lingual fragmentation are consistent with a larger number of subdivisions. The countries that are experiencing a push in opposing directions from these factors are the countries on the off diagonal. It is clear from this graph that all four quadrants have a large number of countries in them, demonstrating that the nonlinearity is an in-sample phenomenon. Table 4 breaks down the set of countries by the four quadrants above and adds the language dimension. The top three rows of the table indicate whether a country's heterogeneity along each dimension is consistent with a higher (+) or lower (−) number of subdivisions. Instead of showing all eight possible scenarios with respect to countries' location relative

TABLE 4: Summary Statistics by Religious, Ethnic, and Lingual Fragmentation Effects

	(1)	(2)	(3)	(4)	(5)
Effect of ethic fragmentation	(+/−)	(−)	(+)	(+)	(−)
Effect of religious fragmentation	(+/−)	(−)	(+)	(−)	(+)
Effect of lingual fragmentation	(+/−)	(−)	(+)	(−)	(+)
Divisions	13.93	13.50	4.00	18.95	11.18
Population	3.18	1.65	0.26	2.16	1.51
Land area	6.85	5.49	0.11	8.12	6.63
Landlocked	0.20	0.14	0.17	0.07	0.25
Coastline	3.84	1.61	0.78	3.39	5.32
Island	0.23	0.05	0.67	0.14	0.18
Harbors	5.75	5.95	3.50	7.67	4.32
Waterways	3.22	2.98	0.02	3.97	1.44
Border length	2.51	2.79	0.21	2.72	3.05
Border number	3.14	3.32	0.83	3.57	3.52
Africa	0.26	0.18	0.17	0.17	0.57
North America	0.14	0.14	0.67	0.12	0.14
South America	0.06	0.23	0.00	0.07	0.00
Asia	0.23	0.36	0.00	0.24	0.11
Oceania	0.08	0.00	0.17	0.02	0.09
Ethnic	0.44	0.61	0.19	0.21	0.74
Language	0.39	0.27	0.58	0.17	0.76
Religion	0.44	0.20	0.57	0.23	0.65
GDP	8.72	7.69	4.17	11.80	4.86
Gini	39.48	44.38	28.90	37.04	43.99
Number of countries	197	22	6	42	44

Note: This table lists summary statistics for all countries in the first column. The last four columns show summary statistics based on the sign of the estimated marginal effects on ethnic, religious, and lingual fragmentation from Model (1) in Table (3).

to the turning points, we focus on the extreme cases. The first column shows the summary statistics for all countries. The second and third columns show the summary statistics for the variables of interest for countries that are experiencing a push in the same direction from the three measures of heterogeneity. Columns four and five show summary statistics for countries that are experiencing a push-and-pull from the different heterogeneity measures.

Column (4) shows the set of countries that experience a push toward more subdivisions from ethnic heterogeneity and a pull toward fewer subdivisions from religious and lingual heterogeneity. Relative to the sample mean, these countries have more subdivisions, are less densely populated, and significantly wealthier than the sample average. The largest share (38 percent) of these countries are on the European continent. The fifth column shows the set of countries that experience a pull toward fewer subdivisions from ethnic heterogeneity and a push toward more subdivisions from religious and lingual heterogeneity. Relative to the sample mean, these countries have slightly fewer subdivisions, are less densely populated, and significantly poorer than the sample average. The largest share of these countries is from Africa.

5. CONCLUSIONS

This paper examines the relative importance of counties' size, natural transportation infrastructure, location, population fractionalization, and level of development in explaining observed variation in the number of first-order political subdivisions across countries. We find statistical evidence suggesting that larger countries that are landlocked have a larger number of neighbors and harbors have a larger number of political subdivisions. We find evidence that island nations and countries with longer borders and more waterways have fewer subdivisions.

The most significant finding is that the number of first-order subdivisions is affected by the degree of heterogeneity of the population. However, the sign of this effect depends on the type of heterogeneity considered and how fragmented the population is along the dimension considered. We show evidence of a statistically significant inverse-U relationship between the number of jurisdictions and a measure of ethnic heterogeneity. In addition, religious and lingual heterogeneity seem to affect the number of subdivisions in a highly nonlinear way, yet opposite in sign to that of ethnic heterogeneity. This finding contradicts what Alesina et al. (2004) observe for school districts across U.S. counties. We argue that different measures of population heterogeneity may work at separate levels of aggregation—states versus counties and cities—and in different directions.

The interesting questions arising from the results in this paper are why we observe the nonlinearity in heterogeneity and why it works in different directions for the different indicators. Although it is beyond the scope of this paper, it seems evident that there is at least one additional effect driving the number of political subdivisions along the heterogeneity dimension. One explanation for the nonlinearity on ethnic fractionalization is that the institutions deciding on the number of first-order subdivisions may not be benevolent social planners, but do exhibit rent maximizing behavior. Alesina and Spolaore (2003) show that these "Leviathan" governments will divide a country into fewer subdivisions than optimal under the central planner regime, which may explain why we observe an inverse-U in ethnic fractionalization.

One potential explanation for the highly significant and unexpected effect of religious and lingual heterogeneity is that in order to prevent secession of minorities, governments design much smaller subdivisions, making it harder and less appealing to secede for any individual nation. The big and positive outliers here are Thailand, Turkey, and Vietnam. The fact that these countries have a large number of subdivisions could be explained by the fact that two of the countries have an active secessionist movement (e.g., the Kurds in Turkey).

REFERENCES

Alesina, A., R. Baqir, and W. Easterly. 1999. "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics*, 114(4), 1243–1284.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg. 2003. "Fractionalization," *Journal of Economic Growth*, 8(2), 155–194.

Alesina, A., C. Hoxby, and R. Baqir. 2004. "Political Jurisdictions in Heterogeneous Communities," *Journal of Political Economy*, 112(2), 348–396.

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

———. 2003. *The Size of Nations*. Cambridge, MA: MIT Press.

———. 2006. "Conflict, defense spending, and the number of nations." *European Economic Review*. 50(1), 91–120.

Alesina, A. and R. Wacziarg. 1998. "Openness, Country Size and Government," *Journal of Public Economics*, 69(3), 305–321.

Berry, S. T. and J. Waldfogel. 1999. "Mergers, Station Entry, and Programming Variety in Radio Broadcasting," *RAND Journal of Economics*, 30(3), 397–420.

Chambers, R. G. 1998. "The Structure of Multiplant Technologies," *American Journal of Agricultural Economics*, 80(4), 839–851.

Coase, R. 1937. "The Nature of the Firm," *Economica*, 4, 386–405.

Gabaix, X. and Y. M. Ioannides. 2004. "The Evolution of City Size Distributions," in V. Henderson and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics*, Vol. 4, Amsterdam: Elsevier, chapter, 53, pp. 2341–2378.

Kim, S. and R. A. Margo. 2004. "Historical Perspectives on U.S. Economic Geography, in V. Henderson and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics*, Vol. 4, Amsterdam: Elsevier, chapter, 66, 2981–3019.

Krugman, P. 1993. "On the Number and Location of Cities," *European Economic Review*, 37(2-3), 293–298.

Lambert, D. 1994. "Zero-Inflated Poisson Regression with an Application to Defects in Manufacturing," *Technometrics*, 24(1), 1–14.

Law, G. 1999. *Administrative Subdivisions of Countries*. Jefferson, NC: McFarland & Company Inc.

Lucas, R. E. 1978. "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, 9(2), 508–523.

World Bank. 2000. "World Development Indicators 2000," Electronic Database: CD ROM.