

112

THE WEALTH AND SIZE OF NATIONS

Donald Wittman
Economics Department
University of California
Santa Cruz, CA 95064

Phone (831) 459-4445
FAX (831) 459-5900
wittman(at)cats.ucsc.edu

ABSTRACT

This paper provides a general theory explaining the geographic and population size and wealth of nations. Successful countries create conditions for high productivity in the economic sphere by enforcing property rights and providing social overhead capital and at the same time minimize political costs by creating a system of rules that reduces influence costs and allows for diverse preferences. Countries also need an effective military apparatus to protect their wealth from predation by other countries. Success in these endeavors may lead to immigration and geographical expansion, while an inability to meet these goals may lead to extensive emigration or breakup of the country. The argument is done within the context of a formal model that integrates spatial political costs with the benefits of spatially determined economic production and the effect of coercive transfers. The analysis is used to provide insight into secessions and mergers of nation states. Several historical events are covered.

JEL#: D, D7, F2

THE WEALTH AND SIZE OF NATIONS

Donald Wittman*

Department of Economics

University of California

Santa Cruz

March 2000

The study of history is predominantly the study of nations: their rise and decline, consolidation and breakup, and their wars of expansion and independence. From Alexander the Great, through the decline and fall of the Roman Empire, to the present events in Hong Kong, Quebec, Kosovo and Palestine, the issues of sovereignty, merger and dissolution are paramount.

This article develops an analytic framework for understanding these historical events. Section A presents a broad discussion of the factors that affect the geographic and population size of nations. Section B develops a formal model integrating spatial political models with spatially determined Cobb-Douglas production functions and military coercion functions to explain the wealth of nations. Section C derives propositions regarding the size of nations. Section D compare my model to previous work. Section E uses the theoretical constructs developed in the previous sections to explain some important historical events. Section F is the conclusion.

A. THE GENERAL THEORY

Individuals acting alone and in groups try to maximize their welfare. Depending on the group's ability to resolve certain public goods problems, the group may be more or less success-

ful in promoting these ends. In turn, these groups interact either cooperatively or coercively with other groups to achieve the goals of their members. An important collection of groups is the nation state. Typically those in power in one nation have very limited power in another nation state. This is what is known as sovereignty. Different nations are more or less successful in resolving political conflict and promoting economic welfare. Nations too interact with other nations to maximize the wealth of their ruling groups. Sometimes this involves cooperation and sometimes this involves coercion. In this paper we show how this competition, cooperation and conflict affect the equilibrium geographic and population size of nations. In particular, we show that the size and wealth of nations depends on the following three elements: (1) the production technology; (2) political transaction costs; and (3) the military coercion technology.

Nations are a nexus of public goods. A wise public policy choice may significantly increase the overall wealth of the citizenry. Successful countries create conditions for high productivity in the economic sphere by enforcing property rights and providing social overhead capital and at the same time minimize political costs by creating a system of rules that reduces influence costs and allows for diverse preferences. Countries also need an effective military apparatus to protect their wealth from predation by other countries. As we will show, success in these endeavors may lead to immigration and geographical expansion, while an inability to meet these goals may lead to extensive emigration or a break up of the country.

Countries will have an incentive to merge (break up) when the value is greater (less) for a unified country than as separate sovereign states.[1] In theory, economic value is unequivocal if those who want union can bribe those who do not want union, but those who do not want union cannot bribe those who want union to change their minds.

Economies of scale and scope are two basic explanations for the size of nations. The costs of administration and policy coordination are unlikely to be proportional to the size of the nation. Different political systems have different economies of scale. Participatory democracy puts severe constraints on the size of the polity. The central command and control apparatus of the communist system enabled the Soviet economy to deal effectively with large-scale physical capital accumulation and the demands of warfare, but less so in an economy based on human capital. Administrative technology also plays a role. Pre-literate societies tend to be less extensive than those that can keep records.

At some point, diseconomies of nation scale arise. The main reason for this appears to be the political costs of integrating people with diverse preferences and skills. Nevertheless, some countries' legal and political systems are better suited for dealing with diverse preferences and hence are more likely to experience significant levels of immigration.

The merger of states reduces inter-state transaction costs but increases intra-state transaction costs.[2] The impetus to economize on transaction costs affects the number and size of nations. The appropriate forum, within a merged nation or across two sovereign nations, for resolving conflict depends on the political institutions available in each sphere and the nature of the potential conflict.

Different national and trans-national political institutions have different transaction costs and thus the comparative advantage of size depends on which political institutions are in place.[3] The existence of the World Court, the willingness of nations to forego war as a means of settling disputes, and other methods of reducing inter-state transaction costs reduce the optimal nation size. Similarly, an international regime of free trade among nations allows for smaller scale political units since economic production is not limited by demand and supply of

domestic markets. As another example, if people in different countries deal with each other mainly through trading companies (as in the past) or through multinational corporations (today), there is less need for other institutional arrangements.

So far I have not mentioned the standard public finance explanation for the size of jurisdictions -- that in the absence of other costs, the extent of the jurisdiction covers the extent of the externality. Since "externalities" can be dealt with across jurisdictions in a Coasean bargain, "externalities" alone cannot explain the size of the political entity.

The circumstances surrounding merger and devolution of countries need not be peaceful or voluntary. The history of nations has in large measure been a history of exploitation by one country of another interrupted by attempts to resist such exploitation.[4]

When the technology of coercion has greater economies of scale we are likely to see mergers of states in order to exploit others more or to be less exploited. When there are lower military economies of scale or, there are great military economies of scale but military might becomes less important for gaining wealth, then there will be more nations and their average size will be smaller.

The effect of scale in warfare during the middle ages has been well articulated by a number of authors (Bean, 1973; McNeil, 1982; Batchelder and Freudenberger, 1983; and Dudley, 1991). In the earlier middle ages, castles and fortified towns dominated the defensive deployment of mobile infantry units in open combat. There was no comparative advantage from centralized coordination. The development of the mobile siege cannon in the beginning of the sixteenth century shifted the advantage from defensive to offensive warfare in Western Europe. Such warfare required more centralization and an increase in the size and scope of nations.

Because the scale effects of military power have been so well studied by the above authors, in this paper, we will, for the most part, put aside a detailed discussion of economies of scale in coercive power.

Exploitation need not result in a loss of sovereignty because there are other means of transferring wealth from one country to another such as tribute, reparations, theft and bribes. That is, given exploitation, the most efficient (i.e, wealth maximizing) methods will generally be invoked. When a robber says your wallet or your life, you hand over the wallet because both of you are likely to be better off by that exchange. By analogy, both countries may be better off if the exploited country provides tribute rather than the exploiting country taking over the reigns of government and extracting wealth directly.

Because, in general, there are no third party enforcers of international agreements, merger and dissolution agreements may not be credible. In turn this means that wealth maximizing changes in the size of countries may not occur. For example, suppose that a merger between Iraq and Kuwait were wealth maximizing, then Saddam Hussein might promise to make the rulers of Kuwait better off if they merged with Iraq. But such promises are not credible; once Saddam Hussein was in control of Kuwait, he could renege on his promises with impunity. On the other side, suppose that maintaining Kuwait's independence is wealth maximizing. A promise by the leaders of Kuwait to indefinitely bribe Iraq so that it will not invade is not credible either because Kuwait might eventually achieve the means to resist such exploitation.

Thus a complete analysis would require first a determination of the wealth maximizing solution and then an investigation whether there were credible methods of enforcing such a solution. Because this paper is already very long and the wealth maximization solution is logically prior, the issue of credible commitments is left to the companion paper, "The Size of Nations:

the Role of Credible Commitments."

B. A FORMAL MODEL: ASSUMPTIONS

We now present a formal model. Of necessity it cannot encompass all of the factors that we have discussed. However, it does meld together three separate intellectual endeavors -- spatial voting models, economic production functions, and anarchic systems -- into a multi-nation analysis. We build on important papers by Alesina and Spolaore (1996, 1997) on the number of countries and by Skaperdas (1992), Grossman and Kim (1995), and Hirshleifer (1995) on anarchic and predatory systems.[5]

We first consider political costs, then economic production, and finally the role of cross-country coercion on the collective welfare of a country's citizens.

B.1. Political Costs

We assume that individuals have single-peaked political preferences along a $[0, 1]$ continuum, with quadratic loss functions. The distribution of most preferred positions, x , is characterized by $F(x)$. $F(1) = X$ is the total population in the world. The population of country i is

represented by: $\int_{x_{i-1}}^{x_i} f(x)dx = F(x_i) - F(x_{i-1})$. That is, country i is composed of individuals with

preferences between x_{i-1} and x_i . Because equilibrium conditions impose contiguous preferences, this assumption could be seen as an implication of the model.

μ_i , a point along the continuum, represents the political position implemented by country i . A person in country i whose most preferred position is at x faces a quadratic political loss of $[\mu_i - x]^2$. The total political loss to all the people in the country is represented by the following expression:

$$(1) - A_i \int_{x_{i-1}}^{x_i} [\mu_i - x]^2 f(x) dx$$

$A_i \geq 1$ represents the political inefficiency of the government -- the larger A_i , the greater the political cost. Some countries' legal and political systems are better suited for dealing with diverse preferences and resolving conflict at low cost. These countries have smaller A_i . For any given level of A_i , the more diverse the preferences, the greater the political costs.

Thus equation 1 represents the political loss from creating one set of rules (public good) for people with diverse preferences.

B.2. Production Function

Land, denoted by y , exists on a $[0, Y]$ continuum. Country i encompasses $[y_{i-1}, y_i)$ where $i = 1, 2, \dots, n$; $y_{i-1} \leq y_i$; $y_0 = 0$; and $y_n = Y$.

Economic output in country i is characterized as a Cobb-Douglas production function, with

land $(y_i - y_{i-1})$ and population $(\int_{x_{i-1}}^{x_i} f(x) dx)$ as inputs, and scale parameter, S :

$$(2) B_i [y_i - y_{i-1}]^c \left[\int_{x_{i-1}}^{x_i} f(x) dx \right]^{S-c}$$

$1 > B_i > 0$ represents the economic efficiency of country i . A country with poorly defined property rights will have a lower B_i and a lower output than otherwise.

We assume that there are economies of scale in production ($1 < S < C + 1$; $C < 1$). In a regime of perfect free trade and perfect enforcement of property rights across countries, economies of nation scale would be one. Because there is neither perfect free trade nor perfect cross-country enforcement of property rights, inter-country transaction costs are greater than intra-country transaction costs, and therefore within country economies of scale are greater than

one. Thus other things being equal, a larger country will be economically more productive. From the opposite perspective, an international regime of free-trade among nations would reduce S toward 1 and allow countries to be smaller (here, we treat S as a parameter).[6]

$$\text{In order for country } i \text{ to be viable, per capita output, } B_i[y_i - y_{i-1}]^c \left[\int_{x_{i-1}}^{x_i} f(x)dx \right]^{S-c} / \left[\int_{x_{i-1}}^{x_i} f(x)dx \right],$$

must be sufficiently large to sustain life (that is, greater than some value, L).

B.3. Coercion

Coercion is an important method for one nation to increase its wealth (at the expense of another nation). Military expenditures enhance a nation's ability either to exploit the weak or to be less exploited by the strong. But more guns also means less butter -- there is a tradeoff between the direct cost of military expenditure and its indirect benefit.

We will now try to capture these ideas in terms of a coercion or threat function. The concept of a coercion function is far less developed than the notion of a economic production function. There is no canonical form that is generally acceptable. Because coercion plays a relatively minor role in our analysis, we will provide a simple model that captures the main issues without unduly complicating the analysis.

Let m_i be country i 's military expenditures.

e_i ($0 < e_i \leq 1$) is a military efficiency parameter (playing the same role as A_i and B_i did earlier); the larger e_i is, the greater bang (both literally and figuratively) the country gets for a dollar of military expenditure. g ($0 < g \leq 1$) is a general technology parameter, which we will discuss at greater length later in the paper.

We assume that each country interacts militarily only with its neighbors, $i-1$ and $i+1$. It is useful to think of the countries being arcs on the circle (with the 0 point being defined as the right boundary of country 1). Then country n is geographically, but not politically, next to country 1. That is, country 1's neighbor on the left, $1-1 = 0$, is country n . Thus m_0 and m_n are alternative ways of representing military expenditures by country n ; m_{n+1} and m_1 are alternative ways of representing military expenditures by country 1.

The amount of income transferred from $i-1$ to i and from $i+1$ to i is the extortion transfer function, or:

$$gT(e_i m_i, e_{i-1} m_{i-1}) + gT(e_i m_i, e_{i+1} m_{i+1})$$

We assume that $T_1 > 0$, $T_2 < 0$, $T_{11} < 0$, $T_{22} > 0$, $T_{12} > 0$, and $T_{11}T_{22} > T_{12}T_{21}$. That is, the greater i 's military effectiveness, $e_i m_i$, the more that country i extorts from her neighboring countries $i-1$ and $i+1$ (or the less that her neighbors extort from i). This effect is subject to decreasing returns from expenditures. $T_{12} > 0$ implies that the marginal benefit of i increasing her military expenditures is greater when $i+1$'s military expenditure is larger than when $i+1$'s military expenditure is smaller. In turn, these assumptions imply that that $T_1(v, z) \leq T_1(z, v)$ for $v > z$ since $T_{11} < 0$ implies $T_1(v, z) < T_1(z, z)$ and $T_{12} > 0$ implies $T_1(z, z) < T_1(z, v)$. [7] A certain amount of symmetry is also assumed. In particular, $T(z, z) = 0$; that is, there are no transfers when both countries are balanced in military power.[8]

Essentially, military power is used to coerce wealth transfers from neighboring countries. The expected return from warfare is the credible demand by the extorting country -- wars need not be fought.

The benefits and costs of extortion to country i are then captured by the following equation:

$$(3) \quad gT(e_i m_i, e_{i-1} m_{i-1}) + gT(e_i m_i, e_{i-1} m_{i+1}) - m_i$$

B.4. Welfare Maximization

Combining equations 1, 2 and 3, the total wealth of country i is the sum of economic production minus political costs plus coerced transfers from other countries (which may be negative) minus military expenditures. This is represented by the following equation:

$$(4) \quad W^i = B_i [y_i - y_{i-1}]^c \left[\int_{x_{i-1}}^{x_i} f(x) dx \right]^{s-c} - A_i \int_{x_{i-1}}^{x_i} [\mu_i - x]^2 f(x) dx + gT(e_i m_i, e_{i-1} m_{i-1}) + gT(e_i m_i, e_{i-1} m_{i+1}) - m_i$$

Our analysis is in terms of a social planner who wants to maximize total welfare, \bar{W} , over all countries, where

$$(5) \quad \bar{W} = \sum_{i=1}^n W^i + VX.$$

The V term stands for the utility to each of the X individuals in the world from being alive. We assume that V is always greater than any political loss, $(x - \mu_i)^2$, that any individual with preference x might face (there are few martyrs in the world).

We employ a social welfare maximization approach because it provides the intuition for understanding the underlying economic forces that determine the size of nations without the need to employ complex bargaining models. Of course, in the real world, there is no social planner. But, as we will argue later, the comparative statics are in the same direction in a world of anarchy.

We assume that the social planner has control over the assignment of land and population and μ_i , but has no control over military expenditures, which will always be in total greater than

0. The reason for this assumption is two-fold. First, there is no such thing as a neutral third-party enforcer. Any third-party enforcer could use its might to extort other countries. Thus countries must rely on their own military power; it is their ultimate credible threat. Second, there are great returns to military expenditures when the other side has very low military expenditures. For example, if country A has zero nuclear weapons and country B has 5 nuclear weapons, the balance of power is greatly weighted toward country B. If country A has 100 nuclear weapons and country B has 105, there is only a slight imbalance in favor of B. This relationship can also be derived from our assumptions as $T_1(v, v) < T_1(v, z) < T_1(z, z)$ for $v > z$. Because detection is never perfect, the dangers from the other side cheating on an arms control agreement is less if both have a reasonable level of defensive capability. Therefore, disarmament agreements will not result in zero military expenditures.

We will initially confine our analysis to interior solutions. Later we will consider how corner solutions alter the results.

C. PROPOSITIONS

C.1. The relative size of two countries

In this section, I show how an exogenous change in one country's economic and political efficiency parameters affects its size and the size of its neighbors (the number of countries remains constant).

The social welfare maximizer chooses that population and land configuration and level of μ_i that maximizes

$$(6) \quad \bar{W} = \sum_{i=1}^N W^i + VX = - \sum_{i=1}^N A_i \int_{x_{i-1}}^{x_i} [\mu_i - x]^2 f(x) dx + \sum_{i=1}^N B_i [y_i - y_{i-1}]^c \left[\int_{x_{i-1}}^{x_i} f(x) dx \right]^{s-c} + VX$$

$$+ \sum_{i=1}^N gT(e_i m_i, e_{i-1} m_{i-1}) + \sum_{i=1}^N gT(e_i m_i, e_{i+1} m_{i+1}) - \sum_{i=1}^N m_i$$

We note that the coercive transfer terms (those with T) will sum to 0; transfers do not involve a net-social cost. We will concentrate on the border between i and $i+1$. A complete analysis would consider all of the partial derivatives.

The first order conditions for an interior maximum are:

$$\frac{\partial \bar{W}}{\partial \mu_i} \equiv \bar{W}_{\mu_i} = -A_i \int_{x_{i-1}}^{x_i} 2[\mu_i - x]f(x)dx = -2A_i[\mu_i \int_{x_{i-1}}^{x_i} f(x)dx - \int_{x_{i-1}}^{x_i} f(x)dx] = 0$$

$$\begin{aligned} \frac{\partial \bar{W}}{\partial x_i} \equiv \bar{W}_{x_i} &= -A_i[\mu_i - x_i]^2 f(x_i) + A_{i+1}[\mu_{i+1} - x_i]^2 f(x_i) \\ &+ B_i[S - C][y_i - y_{i-1}]^C [F(x_i) - F(x_{i-1})]^{S-C-1} f(x_i) \\ &- B_{i+1}[S - C][y_{i+1} - y_i]^C [F(x_{i+1}) - F(x_i)]^{S-C-1} f(x_i) = 0 \end{aligned}$$

$$\frac{\partial \bar{W}}{\partial y_i} \equiv \bar{W}_{y_i} = B_i C [y_i - y_{i-1}]^{C-1} [F(x_i) - F(x_{i-1})]^{S-C} - B_{i+1} C [y_{i+1} - y_i]^{C-1} [F(x_{i+1}) - F(x_i)]^{S-C} = 0$$

The first equation states that country i chooses a political position, μ_i , to coincide with the mean preference of the country i 's population; that is, $\mu_i = \frac{1}{F(x_i) - F(x_{i-1})} \int_{x_{i-1}}^{x_i} x f(x) dx$. The second equation states that population is shifted until marginal productivity gains minus marginal political cost is equated across countries. While the third equation states that land is shifted until marginal productivity of land is equated across countries.

Typically, spatial political models predict that government policy will be at the median, but the mean makes more sense in this case. First, as just shown, when voters have similar quadratic loss functions, the mean preference maximizes welfare. Thus, with suitable transfers, the mean would be preferred by a majority of voters to the median. Second, if the probability of a voter voting for a candidate is a scalar function of the differences in quadratic loss from the candi-

dates' positions, then vote-maximizing candidates will choose the mean. For example, if the policy of candidate j is z_j and the probability that a person whose most preferred position is at x votes for candidate 1 is $P = 1/2 - [x - z_1]^2 + [x - z_2]^2$, then the expected vote for candidate 1 is $\int \left[1/2 - [x - z_1]^2 + [x - z_2]^2 \right] f(x) dx$. Candidate 1 will choose that z_1 that maximizes this integral. First order conditions are $-2 \int (x - z_1) f(x) dx = 0$ or $z_1 = \bar{x}$. Hence, vote maximizing candidates choose the mean policy position. Thus we do not need a social planner to obtain the result that a country's policy position is at mean preference of its citizens. We could have derived it from vote maximization or just assumed the relationship for each country.

Of course, it is possible that a country does not choose the mean as its policy position. The country's political system may be incapable of creating the kinds of trades necessary to reach such a wealth maximizing solution. Or, there may be a corner solution where a subset of the polity has insufficient tradable assets to encourage a policy move in their direction toward the mean (see section C4). Either way, there is a loss of welfare. For reasons of analytic tractability, we have chosen to characterize this loss of welfare in terms of a greater A_i rather than as a movement from the mean.

So far we have viewed the migration of people and the redrawing of borders as being under the control of a social planner. But in the real world there is no social planner; instead countries are to a great extent in a state of anarchy. Nevertheless, the comparative statics should be in the same direction. That is, other things being equal, if there is an increase in the marginal productivity of labor in country i , net immigration to country i tends to increase; equivalently, net emigration from country i tends to decrease. This is what might be called the comparative-statics version of the Coase conjecture -- even when there are positive transaction costs, factors are more likely to go where they are most highly valued, the greater the differential in value. For

example, Eastern bloc countries captured the surplus from emigration by charging either the emigrant (as the Soviet Union did) or the immigrating country (as East Germany charged West Germany). If there were free migration, the individual or the country of immigration would have captured most of the surplus and there would have been more migration. However, the direction of migration is the same, whoever receives the surplus.

Second order conditions:

First looking at terms involving μ_i .

$$\begin{aligned}\bar{W}_{\mu_i\mu_i} &= -A_i[F(x_i) - F(x_{i-1})] < 0 \\ \bar{W}_{\mu_i x_i} &= -\frac{\partial A_i \int_{x_{i-1}}^{x_i} [\mu_i - x]f(x)dx}{\partial x_i} = 0\end{aligned}$$

Since the average deviation from the mean is always zero whatever the size of the country.

$$\bar{W}_{\mu_i y_i} = 0$$

Given the above relationships and the fact that $\bar{W}_{\mu_i A_i}$ and $\bar{W}_{\mu_i B_i}$ also equal zero, we will henceforth be able to ignore second partials involving μ_i and instead concentrate on the remaining second order conditions. That is, we will be able to analyze relations as if there were only two variables -- x_i and y_i .

$$\begin{aligned}\bar{W}_{x_i x_i} &= 2A_i[\mu_i - x_i]f(x_i) - 2A_{i+1}[\mu_{i+1} - x_i]f(x_i) \\ &\quad + B_i[S - C][S - C - 1][y_i - y_{i-1}]^C [F(x_i) - F(x_{i-1})]^{S-C-2} [f(x_i)]^2 \\ &\quad + B_{i+1}[S - C][S - C - 1][y_{i+1} - y_i]^C [F(x_{i+1}) - F(x_i)]^{S-C-2} [f(x_i)]^2 < 0\end{aligned}$$

$$\begin{aligned}\bar{W}_{y_i y_i} &= B_i[C - 1]C[y_i - y_{i-1}]^{C-2} [F(x_i) - F(x_{i-1})]^{S-C} \\ &\quad + B_{i+1}[C - 1]C[y_{i+1} - y_i]^{C-2} [F(x_{i+1}) - F(x_i)]^{S-C} < 0\end{aligned}$$

$$\bar{W}_{x_i y_i} = B_i C [S - C] [y_i - y_{i-1}]^{C-1} [F(x_i) - F(x_{i-1})]^{S-C-1} f(x_i)$$

$$+ B_{i+1}C[S - C][y_{i+1} - y_i]^{C-1}[F(x_{i+1}) - F(x_i)]^{S-C-1}f(x_i) > 0$$

$$H \equiv \bar{W}_{x_i x_i}^i \bar{W}_{y_i y_i}^i - \bar{W}_{x_i y_i}^i \bar{W}_{y_i x_i}^i > 0$$

Thus the hessian of second order conditions is negative definite.

Proposition 1: Given wealth maximization, if country i experiences an exogenous decrease in political or economic efficiency (that is, A_i increases or B_i decreases), then i will experience a decrease in both land size and population size.

Proof:

First we consider the effect of an increase in A_i . Making use of the implicit function theorem, we take the total derivative of the first order conditions:

$$D\bar{W}_{x_i} = \bar{W}_{x_i x_i} \Delta x_i + \bar{W}_{x_i y_i} \Delta y_i + \bar{W}_{x_i A_i} \Delta A_i = 0$$

$$D\bar{W}_{y_i} = \bar{W}_{y_i x_i} \Delta x_i + \bar{W}_{y_i y_i} \Delta y_i + \bar{W}_{y_i A_i} \Delta A_i = 0$$

Solving the above equations, we get:

$$\frac{\Delta x_i}{\Delta A_i} = \frac{[\mu_i - x_i]^2 f(x_i) \bar{W}_{y_i y_i}}{H} < 0$$

$$\frac{\Delta y_i}{\Delta A_i} = - \frac{[\mu_i - x_i]^2 f(x_i) \bar{W}_{x_i y_i}}{H} < 0.$$

It makes more sense to consider an increase in B_i (that is, an increase in economic efficiency).

Again going through a similar process as above, we get:

$$\frac{\Delta x_i}{\Delta B_i} = \frac{-[S - C][y_i - y_{i-1}]^C [F(x_i) - F(x_{i-1})]^{S-C-1} f(x_i) \bar{W}_{y_i y_i} + C[y_i - y_{i-1}]^{C-1} [F(x_i) - F(x_{i-1})]^{S-C} \bar{W}_{x_i y_i}}{H} \leftarrow$$

$$\frac{\Delta y_i}{\Delta B_i} = \frac{[S - C][y_i - y_{i-1}]^C [F(x_i) - F(x_{i-1})]^{S-C-1} f(x_i) \bar{W}_{x_i y_i}^i - C[y_i - y_{i-1}]^{C-1} [F(x_i) - F(x_{i-1})]^{S-C} \bar{W}_{x_i x_i}^i}{H} \leftarrow$$

These latter two equations are both greater than zero since $H > 0$, $\bar{W}_{x_i x_i}^i < 0$, $\bar{W}_{y_i y_i}^i < 0$, and $\bar{W}_{x_i y_i}^i > 0$.

q.e.d.

These results are not surprising. A decrease in political satisfaction will result in emigration to another country. In turn, this reduces the marginal productivity of the other input, in this case land (the same would hold for capital if we considered it explicitly). Similarly, a decrease in economic efficiency reduces the marginal product of both land and labor, which in turn results in a decrease in the geographic and population size of the country.

C.2. Identical countries where the number of countries is endogenous

Suppose that all countries are identical and that f has a uniform distribution. We first find the equilibrium number of countries and then find the effect of a global exogenous change in economic and political efficiency on the number and size of countries.

Given the assumption of a uniform distribution and identical nations, the social planner maximizes the following expression:

$$(7) \quad \bar{W} = \sum_{i=1}^N W^i + VX = -NA \int_0^{1/N} \left[\frac{1}{2N} - x \right]^2 X dx + NB \left[\frac{Y}{N} \right]^C \left[\frac{X}{N} \right]^{S-C} - Nm^* + VX$$

where, m^* is the equilibrium level of military expenditures in a state of anarchy. Because people are uniformly distributed on $[0, 1]$ and there are X people in the world, $f(x) = X$. Each country is identical. Thus, the amount of land and population in each of the N countries is Y/N and X/N , respectively. Looking at the first country, the population is uniformly distributed between 0 and $1/N$; so the mean is $1/2N$. Again, net transfer over all countries are 0. Since

$$-NA \int_0^{1/N} \left[\frac{1}{2N} - x \right]^2 X dx = \frac{NAX}{3} \left[\frac{1}{2N} - x \right]^3 \Big|_0^{1/N} = \frac{NAX}{3} \left[\frac{1}{2N} - \frac{1}{N} \right]^3 - \frac{NAX}{3} \left[\frac{1}{2N} - 0 \right]^3 = -\frac{NAX}{12N^3},$$

the summation can be simplified to the following expression:

$$\bar{W} = -\frac{AX}{12N^2} + B \left[\frac{1}{N} \right]^{S-1} Y^C X^{S-C} - Nm^* + VX = -\frac{AX}{12} N^{-2} + BN^{1-S} Y^C X^{S-C} - Nm^* + VX.$$

Maximizing the above expression with respect to N, the first order conditions for an interior maximum are:[9]

$$\bar{W}_N = \sum_{i=1}^N W_N^i = \frac{AX}{6} N^{-3} + B[1-S]N^{-S} Y^C X^{S-C} - m^* = 0.$$

Second order conditions:

$$\bar{W}_{NN} = \sum_{i=1}^N W_{NN}^i = -\frac{3}{N} \frac{AX}{6} N^{-3} - \frac{S}{N} B[1-S]N^{-S} Y^C X^{S-C} = \left[\frac{S}{N} - \frac{3}{N} \right] \frac{AX}{6N^3} - \frac{Sm^*}{N} < 0$$

The last equality makes use of the first order conditions. The inequality holds because $S < 3$.

Proposition 2: A decrease in political efficiency (increase in A) or a decrease in economic efficiency (decrease in B) will lead to more countries (larger N) of smaller size.

Proof:

Taking the total derivative of the first order conditions with respect to a change in A, we get:

$$d\bar{W}_N = \bar{W}_{NN} \Delta N + \bar{W}_{NA} \Delta A = \bar{W}_{NN} \Delta N + \frac{X}{6N^3} \Delta A = 0$$

Equivalently,

$$\frac{\Delta N}{\Delta A} = \frac{-\bar{W}_{NA}}{\bar{W}_{NN}} = \frac{-X/6N^3}{\bar{W}_{NN}} > 0$$

That is, the greater the political inefficiency, the greater the number of countries.

We next consider the effect on N of a decrease in B (i.e, a decrease in economic efficiency).

Again taking total derivatives we get:

$$\frac{\Delta N}{-\Delta B} = \frac{\bar{W}_{NB}}{\bar{W}_{NN}} = \frac{[1 - S]N^{-S}Y^C X^{S-C}}{\bar{W}_{NN}} > 0$$

The last inequality holds because $S > 1$.

Thus a global decrease in economic efficiency leads to a greater number of countries.

q.e.d.

Looking at Proposition 2 from the opposite point of view, the greater overall political or economic efficiency, the fewer the number of countries. These results reinforce the conclusions of the previous section.

We next consider the role of military expenditures on the optimal size of countries.

Proposition 3: If m^* increases, the number of countries decreases and the average size increases.

Proof:

Again making use of the first order conditions,

$$\frac{\Delta N}{\Delta m^*} = \frac{-\bar{W}_{Nm^*}}{\bar{W}_{NN}} = \frac{1}{\bar{W}_{NN}} < 0$$

q.e.d.

Countries can save on military costs by merging since they no longer have to defend against each other. Of course, they face increased internal political costs from unification.

This proposition suggests that the potential for extortion will result in the average size of countries being above the wealth maximizing size even if the relative size of one country vis a

vis another is not affected by the potential for extortion and even if the transfer between any two countries is 0. Thus as military might becomes relatively less important, the size of nations will decrease (this will be considered in greater detail in the following section)..

C.3. Military expenditures

We now consider the equilibrium military expenditures. Again, the analysis focuses on interior conditions. For heuristic reasons we will only consider three countries. If country i maximizes W^i with respect to its military expenditures, m_i , then the first order interior conditions for each country are:

$$(8) \quad W_{m_1}^1 = e_1 g T_1(e_1 m_1, e_3 m_3) + e_1 g T_1(e_1 m_1, e_2 m_2) - 1 = 0$$

$$(9) \quad W_{m_2}^2 = e_2 g T_1(e_2 m_2, e_3 m_3) + e_2 g T_1(e_2 m_2, e_1 m_1) - 1 = 0$$

$$(10) \quad W_{m_3}^3 = e_3 g T_1(e_3 m_3, e_1 m_1) + e_3 g T_1(e_3 m_3, e_2 m_2) - 1 = 0$$

Proposition 4: Assume that there are no coalitions. If $e_i = e$ for $i = 1, 2, 3$, then $m_1 = m_2 = m_3$ and there are no transfers. That is, the only equilibrium that exists is one in which all military expenditures are equal.

Proof:

Since $e_i = e$, the following equalities can be derived from the first order conditions:

$$(11) \quad T_1(em_1, em_3) + T_1(em_1, em_2)$$

$$(12) = T_1(em_2, em_3) + T_1(em_2, em_1)$$

$$(13) = T_1(em_3, em_1) + T_1(em_3, em_2)$$

Suppose contrary to the proposition that $m_3 > m_2 > m_1$. Then the second term in (13) is less than or equal to the first term in (12) since we have assumed that $T_1(v, z) \leq T_1(z, v)$ for $v > z$.

The first term in (12) is strictly less than the first term in (11) since we have assumed that $T_{11} < 0$. Therefore the second term in (13) is strictly less than the first term in (11). By a similar logic, one can show that the first term in (13) is strictly less than the second term in (11). But these two inequalities contradict the equality between (11) and (13). Therefore it cannot be that $m_3 > m_2 > m_1$. A similar argument holds for other orderings of the m_i where at least one of the inequalities holds strictly. A simple investigation shows that the equalities between (11), (12) and (13) hold if $m_3 = m_2 = m_1$.

q.e.d.

We next turn our attention to the case where $e_3 > e_2 > e_1$.

Proposition 5: If $e_3 > e_2 > e_1$, then $e_3m_3 > e_2m_2 > e_1m_1$.

Proof:

The assumption that $e_3 > e_2 > e_1$ along with the first order conditions implies:

$$(14) \quad T_1(e_1m_1, e_3m_3) + T_1(e_1m_1, e_2m_2)$$

$$(15) > T_1(e_2m_2, e_3m_3) + T_1(e_2m_2, e_1m_1)$$

$$(16) > T_1(e_3m_3, e_1m_1) + T_1(e_3m_3, e_2m_2)$$

Suppose contrary to our proposition that $e_1m_1 > e_2m_2 > e_3m_3$. Then the first term in (14) is less than the first term in (15) since $T_{11} < 0$ by assumption. The second term in (14) is less than or equal to the second term in (15) since by assumption $T_1(v, z) \leq T_1(z, v)$ for $v > z$. Hence (14) $<$ (15), contrary to what was stated above. Therefore $e_3m_3 < e_2m_2 < e_1m_1$ cannot hold.

In a similar manner, other orderings besides $e_3m_3 > e_2m_2 > e_1m_1$ can also be shown to imply a contradiction. Only if $e_3m_3 > e_2m_2 > e_1m_1$ will (14), (15) and (16) be satisfied.

q.e.d.

Thus militarily more efficient countries will extort more from other countries. It can also be shown that militarily more efficient countries have greater military expenditures.

We next consider the effect of an increase in g on the equilibrium amount of military expenditures, and in turn, the effect on the number of countries.

Proposition 6: If all countries are identical, an increase in g will result in a decrease in N .

Proof: We first show that an increase in g will result in an increase in m_i for all i .

From 8, 9 and 10, we make use of the implicit function theorem to derive the following:

$$(17) \quad dW_{m_1}^1 = W_{m_1 m_1}^1 \Delta m_1 + W_{m_1 m_2}^1 \Delta m_2 + W_{m_1 m_3}^1 \Delta m_3 + W_{m_1 g}^1 \Delta g = 0$$

$$(18) \quad dW_{m_2}^2 = W_{m_2 m_1}^2 \Delta m_1 + W_{m_2 m_2}^2 \Delta m_2 + W_{m_2 m_3}^2 \Delta m_3 + W_{m_2 g}^2 \Delta g = 0$$

$$(19) \quad dW_{m_3}^3 = W_{m_3 m_1}^3 \Delta m_1 + W_{m_3 m_2}^3 \Delta m_2 + W_{m_3 m_3}^3 \Delta m_3 + W_{m_3 g}^3 \Delta g = 0$$

Since $e_i = e$ and $m_i = m^*$ for all i ,

$$W_{m_i g}^i = e_i T_1(e_i m_i, e_j m_j) + e_i T_1(e_i m_i, e_2 m_2) = 1/g = W_{m_2 g}^2 = W_{m_3 g}^3$$

and $e_i T_1(e_i m_i, e_j m_j) = e T_1(em^*, em^*)$ for all i and j .

Thus 17-19 can be rewritten as:

$$e^6 g^3 \begin{bmatrix} 2T_{11} & T_{12} & T_{12} \\ T_{12} & 2T_{11} & T_{12} \\ T_{12} & T_{12} & 2T_{11} \end{bmatrix} \begin{bmatrix} \Delta m_1 \\ \Delta m_2 \\ \Delta m_3 \end{bmatrix} = \begin{bmatrix} -g^{-1} \\ -g^{-1} \\ -g^{-1} \end{bmatrix}$$

Denoting the terms in the bracket on the left side of the equality by H , H is negative definite by assumption. Solving for the effect of a change in g on m_i , we get:

$$\frac{\Delta m_i}{\Delta g} = - \frac{e^6 g^2 [4(T_{11})^2 - (T_{12})^2 - 2T_{12}T_{11} + (T_{12})^2 + (T_{12})^2 - 2T_{11}T_{12}]}{H} = - \frac{e^6 g^2 [4(T_{11})^2 - 4T_{12}T_{11} + (T_{12})^2]}{H}$$

The expression is greater than 0 because H is negative and the bracketed term is positive since

$T_{12} > 0$ and $T_{11} < 0$ by assumption.

Thus an increase in g results in an increase in m_i . Combining these results with Proposition 3, we get the result that an increase in g results in a decrease in N .

q.e.d.

g is a general military effectiveness variable that depends on both military technology and the spoils of war. Warfare may result in destruction of the object desired by the invading country. Under such circumstances, the threat of war is less credible and therefore the ability to extort wealth is also reduced.[10] In general, unmined minerals and agricultural land are more likely to survive a war than physical capital, and it is still harder to expropriate human capital than physical capital. So countries whose wealth is in the ground tend to be more vulnerable to extortion. The more important this wealth is, the larger g is and the larger the average size of countries.

C.4. Corner solutions

Our analysis has focused on interior solutions; but corner solutions are always possible. In such cases, the anarchic solution is likely to greatly diverge from the social welfare solution. Consider the case where the political cost of integration of a people into any existing country is greater than their marginal productivity minus their minimal daily requirement for existence (which is L per person). This set of people have nothing of value (except their humanity) to offer to others. A social welfare maximizer would either (A) create a new country with sufficient land to produce at least L per person or (B) add these people to an existing country. Because some country is always viable and there are economies of scale, the latter is always an option. Since we have assumed that the value of life, V , is greater than any political costs, a social welfare maximizer would always prefer that these people exist. But in the absence of a social planner, individual countries are unlikely to value these people, qua people, and would prefer to exterminate them rather than have them part of the polity.

This situation is most likely to occur when one country is technologically very far behind other countries. The backward country is then likely to be subject to extortion. Without the productive wherewithal, indigenous people of the backward country will not be able to pay tribute as a substitute for giving up the land. And when the marginal product net of minimum daily requirements is less than the cost of political integration, the extorting country may prefer extermination of the native population. V is not a tradable good.

Alternatively, the imperial leaders may allow the indigenous people to live but not respond to their political wishes. That is, μ would be at the mean preference of the colonial power rather than at the mean preference over the entire population. If the indigenous people had sufficient tradable assets, they would be able to move the policy toward their own interests and the mean overall, but they do not have such tradable assets either because they are not sufficiently productive or because their assets were taken away.

This seems to characterize Europe's policies toward their colonies in the 19th century and China's policy toward Tibet today. The imperial country's need for labor determines its strategy toward the colonial subjects. If its need for labor is small, it will either engage in a policy of genocide or encourage migration. If it needs labor, it will keep payments to the indigenous people to a minimum and extract all tradable surplus. As a consequence, there will be little integration.

D. COMPARISON TO OTHER MODELS

Now that we have specified the model, it is useful to compare this work to others.

In his seminal article, Friedman (1977) viewed the state as being embodied in a ruler who maximizes tax revenues (net of administrative costs).[11] He showed how different sizes and shapes of countries allow for differing levels of tax collection. For example, as labor becomes

more productive, the state will want to increase its tax on labor, but this is limited by the possibility of migration. Therefore, the state will want to enlarge its size and erect linguistic and/or physical barriers (such as the "Iron Curtain") to increase the cost of migration and reduce the competition from other countries. Since trade increases the productivity of labor, Friedman also argued (contrary to my results) that increased trade leads to larger nations in order for their rulers to capture greater monopoly rents.

There are significant differences between Friedman's approach and mine. In Friedman's analysis, net tax receipts are maximized while in my model wealth is maximized.[12] Thus in his model, the outcome depends on net tax revenues to the rulers and not on the costs that might fall on the subjects. Consequently, there is no resistance by the subjects to costly attempts at monopolization by the ruler. However, if Friedman were correct about the exploitation, then people in both countries should resist the merger of two otherwise similar countries because merger would just result in their greater exploitation. On the other hand, if one country could provide a more efficient set of rules and credibly less exploitation, then, other things being equal, that side would gain territory and citizens at the expense of the other side. That is, the subjects will fight harder if they can credibly end up with more afterwards. This may explain why, in fights between democracies and autocracies, democracies have tended to win. The citizen soldiers of democracies have more to lose if the autocracy wins than the citizen soldiers of autocracies lose if the democracy wins (see Lake, 1992, for further discussion and evidence). So even if the rulers have different goals, the outcome will be based on wealth maximization.

Alesina and Spolaore (1996, 1997) have a political cost function, but neither a production function nor an extortion function. Their political cost function differs from mine in a number of ways. They assume that voters have linear, not quadratic, loss functions, that preferences are

uniformly distributed, and that μ is the most preferred position of the median voter, not the most preferred position of the mean voter. In their model all countries are identical and have the same size. In their model behavior is based on averages -- citizens are treated equally; in my model welfare maximization tends to take place so that payments are more likely to depend on marginal contributions. More important, in their model individuals are tied to the land -- there is no migration; in my model, independent migration is possible.

Schmidtchen (1994) and Schmidt-Trenz et. al. (1994) have a series of papers on the optimal size of nations that is based on Buchanan's constitutional theory and the theory of clubs. Unlike the present paper, they do not have a spatial model of political cost nor an explicit production function.[13] Josselin and Marciano (1998) consider the boundary between one state and the periphery. Cost increases with the size of the country, but otherwise they have an entirely different approach.

We have already mentioned the many authors who have discussed scale economies and diseconomies in warfare. Scale economies are also an important part of the new economic geography (see, for example, Fujita, Krugman and Venables, 1999). However, these models do not explain nation size. For example, economic geography might be able to explain why Seattle is a locus in the software industry, but not why Seattle is part of the United States and not Canada.

No economic discussion of nations should neglect the work of North and Olson. North and Thomas (1973) and North (1981) explain why certain states grow economically, not their consolidation and breakup. They argue that not all systems are wealth maximizing. While we incorporate a wealth maximizing explanation, as should be clear from our model, we do not assume that all nations maximize wealth -- just that those countries whose institutional arrangements are better at maximizing wealth are more likely to takeover countries whose institutions are less suc-

successful at wealth maximization rather than the reverse. North and Thomas emphasize the role of property rights; those nations that have a strong system of property rights will grow and prosper. In our model, property rights are only part of the picture.

Olson (1982) also devotes his analysis to economic growth rather than population and geographic growth.[14] In his work, a primary cause of economic stagnation is the power of non-inclusive pressure groups whose welfare is reduced when the economy grows. These groups have sufficient power to prevent growth but not enough power to allow growth and skim its surplus. Again my explanation need not be in opposition. A political system that cannot figure out how to make interest groups benefit from growth will be less powerful economically and possibly militarily (unless the military itself is a successful pressure group that rewards the most successful military innovations). As a result it may be a candidate for takeover. On the other hand, there are differences between his approach and mine. Olson's model has an ontological destructiveness of interest groups so that older political units with older pressure groups tend to grow more slowly. I have a different perspective -- that certain institutional arrangements are more productive under certain circumstances and less under others.

We have considered scale economies throughout the paper. Scale economies are also an important part of the new economic geography (see, for example, Fujita, Krugman and Venable 1999). However, these models do not explain nation size. For example, economic geography might be able to explain why Seattle is a locus in the software industry, but not why Seattle is part of the United States and not Canada.

E. HISTORICAL EXAMPLES

The ideas developed in the previous sections will now be used to provide insight into the history of nations. The empirical evidence will rely on dyads -- two countries or one country in two different time periods which are otherwise similar except for a difference in one of the exogenous variables that we have identified.

D.1. The Roman Republic versus the Roman Empire: the role of A_i

In this subsection, we consider the enduring symbol of imperialism -- the Roman Empire. We show that military strength alone cannot explain whether tribute or territory is given to the victor. A comparison of the beginning of the Roman Empire under Augustus (43 B.C.E. - 14 A.C.E) to the Roman Republic 150 years earlier elucidates the political as opposed to the merely military aspects of empire.

The greatest expansion and consolidation of the Roman Empire took place under Augustus. "[The Augustan political system] was one of the most artful blends of old theories and new principles ever conceived by any statesman. Its satisfaction of the needs and aspirations of the important elements in the Mediterranean world underlay two centuries of peace in the Roman Empire" (Starr, 1983, 554). Although the local populations did not accept Rome with open arms, Roman control created numerous benefits for them. The Romans provided an administrative bureaucracy and laws without extorting too much from the provinces. The following quote from Doyle (1986, 95-96) is instructive. "By removing control of the state from the city of Rome -- from its Senatorial oligarchy, equestrian financiers, and popular assembly -- Augustus established a bureaucracy that put his, its own, and the Empire's interest before any specifically Roman demands. In the absence of such an imperial bureaucracy, the periphery will be ravaged, the metropole divided and both destroyed."

Roman legionnaires also protected the local population from internal disorder (within the empire) and external threat. A permanent navy eliminated piracy. Peace, low taxes, and a sound imperial currency allowed trade to flourish and investments to be made (high B_1). The Roman empire went into decline in later years when internal squabbling and lower productivity in the center lead to greater exploitation of the periphery.

Approximately 150 years before the reign of Augustus, the Roman Republic had twice shown its military might in the eastern Mediterranean, yet on both occasions had chosen tribute rather than territory. In the Second Macedonian War (200-196 B.C.E.), Rhodes and Pergamum enlisted Rome's support against Philip V of Macedonia. Phillip had to make peace in which he promised to stay out of Greece, pay an indemnity of 1000 talents and yield all but five warships. But he was left in control of Macedonia, and Rome extracted no territory. In the war with Antiochus III (192-88 B.C.E.), Pergamum again enlisted the aid of Rome. In the ensuing peace, Antiochus yielded all but ten of his warships and elephants, paid an indemnity of 15,000 talents, and relinquished all Asia minor to Rhodes and Pergamum. Except for a few islands off the Balkans, the Roman Republic had not annexed anything in the East. Why? Internal political weakness and a comparative disadvantage with the politically more advanced Hellenistic states warranted independence for the eastern Mediterranean. Indeed, even in later times, when they came under the control of Rome, the nature of Roman involvement in the Hellenistic states was less than it was in Gaul, for example.

D.2. Kuwait versus Singapore: the role of g and e_1

At the opposite end of the continuum from empires are city states and small countries.

We start with a question. How can Singapore be economically viable? Free trade in a world economy allows economic specialization to take place independent of the political unit.

Therefore, small states can be economically viable, especially if they have access to major trading routes. On the political side, ethnic differences between Singapore and Malaysia would make the cost of consolidation high and are the reason for the breakup in the first place.[15]

Nevertheless, small states are always in danger of being exploited by their bigger neighbors so that, at a minimum, their surplus value is extracted. Singapore is at very low risk however because its wealth is in human capital which is relatively hard to extract; that is, any extortion threat is less credible since the threatening country would end up with very little if it had to carry through with its threat and capture Singapore (g is low).

In contrast, most of Kuwait's wealth is in in the ground; that is, g is very high. Kuwait is also more vulnerable on other grounds. Relative to Iraq, Kuwait's military effectiveness (e_k) is very low. Kuwait remains independent and more or less immune from a takeover because the value of its oil to the United States and Europe is greater than it is to Iraq (it is worth more to the consumers to prevent a monopoly than to the producers to create one). And therefore the consumers (US and Europe) protect Kuwait.

Hong Kong is an interesting middle possibility. Its wealth is again in human capital and therefore difficult to expropriate. While re-integration of Hong-Kong into China was not desired by most of the city's residents, the re-integration may be wealth maximizing if it invigorates the rest of the Chinese economy. This appears to be the case, as even before 1997, the areas closest to Hong Kong were the most economically successful provinces and many of Hong-Kong businessmen had already benefited from close ties with the PRC.

D.3. Athenian democracy: the failure of integration (large A_i)

Ancient Athens is considered the source of democracy, but for most of its existence it remained a city state. It had a dismal record as an imperial power. Why? Ancient Sparta was a

military machine, yet it too had a short-lived and unsuccessful existence as an imperial nation.

Why?

The Classical Period in Greece (750-350 B.C.E) displayed inter-city warfare and coalitions, but no long lasting large federation until the reign of King Philip of Macedonia. This was a period of great innovation in domestic political structure and considerable international instability. Our previous analysis suggests that these two should go hand in hand since changes in political productivity change the optimal size of countries. Even in classical times, the waxing and waning of the various states filled volumes. I will briefly discuss two periods: Pericles and Athens (461-429 B.C.E.); and Spartan hegemony (404-371 B.C.E.).

The evolution of democracy was intimately connected with the increasing imperialism of Athens. On the one hand, imperialism changed the relative wealth and power of citizens within Athens. The commercial elements benefited from the improved trade and were important backers of the democratic assembly over the aristocratic Aeropagus. At the same time, democracy made Athens more effective in exerting its influence abroad. For example, citizens were paid for their military service to the state.[16] Ultimately, the democratic system also limited the ability of Athens to rule her territories effectively. Direct democracy with citizenship only allocated to those who were born of Athenian citizens had made Athens powerful but was an inappropriate structure for a larger state. The other cities were exploited and there were repeated revolts. When the Peloponnesian War broke out, Sparta could stand before Greece as the champion of liberty.

After the Peloponnesian War, Sparta maintained hegemony over the Greek States. Even in Athens, Sparta installed a 30 member oligarchy. However the disjunction between military power and political capability was very apparent. Thucydides (1.77) had been very doubtful that the Spartans could adapt their narrow ways to foreign rule, while Aristotle argued that the Spar-

tans "collapsed as soon as they acquired an empire. They ... had never accustomed themselves to any discipline other and better than that of war." The Spartans were alternately ruthless and ineffective, and, in later years, corrupt. Thebes easily created a federation against Sparta and established its own hegemony in 371 B.C.E.

D.4. France and Poland in the 19th Century: the role of administrative advantage

While not denying the role of the military in shaping the political map, my analysis places military power in a secondary position.

Military victory need not lead to territorial changes. For example, after Napoleon's defeat and his banishment to Elba, The Congress of Vienna extracted virtually no European territorial concessions from France although the allied armies had ended the war at the gates of Paris.[17] Even after Napoleon's return and subsequent defeat at Waterloo, territorial concessions were minimal (the boundaries of France remained virtually unchanged from those at the start of the French Revolution); instead, more severe indemnities were to be paid by the vanquished nation. The end of the Franco-German War again found Paris surrounded, yet territorial concessions were limited to Alsace and Lorraine and control over the Papal territory. Instead of more territory, Germany received 200 million sterling. A trade had been made. If they were rational, there was no other bargain (for example, France giving up more territory, but paying less indemnity) that would have made both sides better off.

A comparison of 19th century Poland and France is instructive. The flat topography of Poland has always been used as a reason for its easy subjugation, but the partition and complete disappearance of Poland in the latter part of the 19th century was also due to its internal weaknesses.[18] Polish nobility had been a very powerful force in maintaining the status quo; any one could veto legislation and dissolve the assembly. This system protected minorities,

including ethnic minorities, but it had its drawbacks. The central government had neither substantial revenues, nor a well equipped standing army, nor a bureaucracy to do the work of a state. Poland did not have an effective political apparatus and was incorporated by her neighbors. In contrast, France maintained most of her territorial integrity, despite having much of her territory occupied in three separate wars. An important difference between the two countries was that, with the exception of those times of internal strife, France had an effective national apparatus; Poland never had much of a national government, at all.[19]

D.5. Eastern Europe and the role of exploitation

A military machine is costly to any society unless it is able to extort more from other nations than it costs to maintain (that is, $T_i(e_i m_i, e_{i-1} m_{i-1}) + T_i(e_i m_i, e_{i+1} m_{i+1}) > m_i$). At the end of World War II, Soviet forces in Eastern Europe were able to dismantle the remains of East Germany's industrial base. The Soviet Union also took industrial equipment from other satellite countries. After that, the Eastern European satellites took on the role of buffer states. But as the stalemate of the cold war went on, it became clear that the military need for buffer states no longer existed. The economic and political "integration" of Eastern Europe with the Soviet Union was a negative-sum game. The communist system was ill suited for modern societies. There were immense difficulties of coordinating a complex society from the center. The political apparatus of the Communist Party was too ossified to deal with either political or economic problems within individual countries, and, with the possible exception of military matters, was incapable of dealing effectively across countries. The coordination of the economies via Comecon was unwieldy. And the centralized and externally imposed political structure was not responsive to local political needs. The economic and political failure of the communist political apparatus became obvious when one saw the thriving economies and polities of Western Europe.

From the Soviet side, the cost of control was greater than any benefits it was receiving. The excessive drain on a weakened economy of a large military apparatus could no longer be justified. Similar forces existed within the Soviet Union. The collapse of the Communist party as a system of political and economic control meant a collapse at the center. There were no centralized institutions, except the army, to administer over the USSR. Most of the republics saw that their wealth would increase by being more independent, and the Union of Soviet Socialist Republics dissolved soon afterwards.

D.6 Yugoslavia: when μ_i is not at the mean

After the death of Tito, the Serbs attempted to increase their own welfare by shifting the policy of the government to the mean of the Serbian position away from the mean of the country as a whole. Not surprisingly, this encouraged the Slovenians, Croatians, Bosnian Muslims and Kosovar Muslims to seek independence. The result was a break-up of Yugoslavia. During the wars in Bosnia and Kosovo, the the Serbs tried to get land for free by ethnic cleansing. Thus the recent history of this area can be summarized as a failed attempt by the Serbs (and Croats) to increase their own wealth at the expense of others.

Note this history can be explained in terms of groups trying to gain economic and political advantage. There is no need to rely on arguments of deep-seated ethnic hatreds that go back for nearly 1000 years.

F. CONCLUDING REMARKS

It is hard to condense the history and theory of the size and wealth of nations into a short paper. Even restricting the analysis to a formal model provides many disparate components. Therefore, I will merely mention some of the important comparative static results.

Nations that can better integrate people with diverse preferences and that can promote property rights will tend to be wealthier with immigration, capital inflows and a larger geographic area.

The number of small states increases as free trade across countries increases. For similar reasons, when the threat of war decreases, the the size of nations also decreases. This has already occurred in the former Soviet Union, and it may be the future for other countries as well.[20]

As land per se becomes less important relative to capital and labor in producing wealth, adjustments will tend to be through changes in capital and migration rather than in land. The value of the military in increasing the nation's wealth will also be reduced as theft and extortion become more difficult. Because mineral wealth is relatively impermeable to war, it will continue to be the major source of conflict over territory.

The analysis need not be restricted to the nation state. Many of the insights provided here can be applied to cities, religions, tribes and clans.[21] For example, Sahlins (1968) reports on a set of African tribes that limit their membership to 50 people. When a tribe gets larger than that, new tribes are created even though there may be tribal warfare. The reported reason for this division is that the political structure is incapable of handling larger entities.

Finally, the theoretical synthesis of the spatial political model with an economic production function can serve as the foundation of future research in numerous other areas, including the theory of the firm, federalism, and the nature of economic development.

FOOTNOTES

* I would like to thank David Friedman, Jack Hirshleifer, Dieter Schmidtchen, Charles Tilly, participants at the Villa Colombella Seminar, Rome, 1996, my colleagues at UCSC and CES, and the referees for valuable comments. I would also like to dedicate this paper to Mancur Olson who provided many comments shortly before his untimely death.

[1] The analysis does not require nations to be at the efficiency frontier -- only that very large gains to consolidation (being separate) are over time likely to lead to unification (devolution).

[2] For a discussion of transaction costs tradeoff in the context of a federal system, see Ellingsen (1998). Yarbrough and Yarbrough (1994) view the difficulty of enforcing cross-border contracts as a key reason for expansion of borders.

[3] In order to make the analysis more manageable, I will not consider the endogeneity of political structure.

[4] Although each country tends to maximize its own wealth, wealth is not maximized for the system as a whole. Scarce resources are used in coercing and resisting coercion. War is an extreme form of transaction costs.

[5] In turn these author's built on earlier works. See Blum and Dudley (1991) for a spatial model of the state and Bush (1972) for a model of the predatory state.

[6] In the model there is only one output and, consequently, no international trade in goods (of course there is "trade" in factors of production when people emigrate and boundaries are redrawn). However, one can conceive of there being international trade with output being the post-trade result. We would still see movement of the factors of production from one country to another just as we see the movement of factors of production from one firm to another within a

country.

[7] These assumptions capture the anarchy of Hirshleifer's 1995 model, the role of extortion in Grossman and Kim (1995), and reflect the approach used by Skaperdas (1992). A major difference is that Skaperdas assumes an S-shaped curve while we have a concave function.

Further elaborations of the extortion model are possible. E.g., Land size and population might be inputs into the extortion function. An extremely small geographical territory may be hard to defend whatever the population. However, for any given population, after some point, it becomes more difficult to defend greater territory.

[8] This last assumption is not necessary for analyzing the effect of coercion. I make this assumption so that I can focus this paper on the other two equations.

[9] We ignore the problem that N may not be an integer.

[10] Since wars are costly to both sides, when there is perfect information, the only reason why a country would go to war where there are no spoils is to establish a reputation. Even if we allow for reputation effects, the cost of establishing a reputation is not zero, so that threats are less effective if there are smaller rewards to the party threatening.

[11] Tilly (1990) has a more nuanced theory of the state where the sovereign maximizes the collection of receipts via war and capital.

[12] Clearly, a more productive society would enable the ruler to collect more taxes. Friedman ignored such issues and instead focused on tax collection.

[13] For earlier work along these lines see Buchanan and Faith (1987). Bolton and Roland (1997) also employ a non-spatial model where the issue is income redistribution.

[14] Kennedy (1987) covers some of the same themes.

[15] A different political system might have been able to integrate these different ethnic groups more successfully, but possibly at the cost of lower economic production.

[16] "When navel power relies on oared galleys, a democratic navy is inherently superior to a nondemocratic navy since the latter is likely to employ slave rowers who in the heat of battle cannot be called upon to fight." Doyle, 1986: 67.

[17] Territorial loss may have impelled greater resistance by the French (why this might be so should be answered within the framework presented earlier). "Balance of power " is the usual explanation for maintaining territorial integrity, but as we will see, balance of power was not invoked with regard to Poland.

[18] A duchy remained under the control of Russia, but it did not maintain the name of Poland.

[19] Poland had different ethnic groups, but one could argue that Bretagne should have gone to England.

[20] Similar forces are at work in China. Power is shifting from the center to the provinces.

[21] The ideas of this paper are complementary to Tiebout models that explain city size.