Origins and Persistence of Economic Inequality

Carles Boix

Department of Politics and Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, New Jersey 08540; email: cboix@princeton.edu

Key Words
development, state, institutions, democracy

Abstract
After reviewing the current literature on the causes of economic inequality, the article models the historical emergence of inequality as the result of a key technological change (i.e., the adoption of agriculture) that widened income differentials and led to the construction of state institutions, which shaped (depending on their particular nature, more or less authoritarian) the final distribution of economic assets within and across different societies. The article then explores the evolution of inequality in societies already endowed with state structures: A stream of biased technological shocks happens randomly and the “decisive” voter (who differs across political regimes) accepts or blocks them as a function of their effect on her net income. The decisive voter’s response determines the overall distribution of income. The model is employed to give a coherent account of some broad historical trends in the evolution of income inequality.
INTRODUCTION

Inequality—its causes as well as its political and social consequences—has been a long-standing concern in political economy. Both Aristotle and Machiavelli associated a particular distribution of wealth with the nature of political institutions. Rousseau devoted his whole Discourse on the Origin and Basis of Inequality among Men to explore the causes (the need for social recognition and the exploitation of man by his own kind) behind the destruction of the moral and political equality that he ascribed to mankind in its infancy.

By comparison with those classic endeavors, contemporary research on the roots of inequality remains rather limited. In recent years, economists have made considerable efforts to investigate the effects of inequality on growth and development (Atkinson & Bourguignon 2000, Kahhat 2007). In political science, there is now a burgeoning literature on its impact on political institutions and conflict (Boix 2003, 2008; Acemoglu & Robinson 2006; Geddes 2007). However, very little theoretical work has been done on the underlying causes of the evolution of inequality. Most contributions are strict accounting exercises or limit themselves to economic variables such as imperfect credit markets, which are in turn left unexplained.

Moreover, probably owing to the scarcity of data, those researchers who have worked on the evolution of income distribution have exclusively focused on its evolution in advanced countries in the past three decades (Gottschalk & Smeeding 1997; an exception is Rogowski & McRae 2008). Yet any reasonable inquiry about the origins and varying persistence of inequality should tackle a much broader temporal span and, in particular, should account for the following set of stylized facts. First, primitive, stateless communities tended to display relatively equal distributions of income and wealth. Although they had some internal heterogeneity in status and prestige as well as some differential access to a few goods (such as sexual partners), their overall level of social and economic stratification was limited and they had similar interpersonal patterns of consumption and individual welfare. Second, that relative equality gives way to much wider distributions of income and wealth—in terms of patterns of habitation, accumulation of valuable assets, and even health and height (Boix & Rosenbluth 2006)—after the agricultural revolution and the formation of states. Third, the extent of social and economic inequality in agricultural societies has varied considerably across areas and historical periods. Finally, income inequality has trended downward, albeit with a few national exceptions and temporal reversions, in the 150 years since the spread of industrial technologies.

Given the incipient state of the literature on the causes of wealth and income inequality, the first part of this article reviews the existing theoretical and empirical research on that question. The remainder of the article then offers a possible analytical strategy that draws on both economic and political approaches to account for the origins and varying persistence of inequality. The second part sketches a model to explain the transition from relatively equal primitive communities with no full-fledged state institutions to human societies with stable political institutions and various degrees of economic inequality. According to that account, the adoption of agriculture both widened income differentials and led to the construction of state institutions, which then shaped (depending on their particular nature, monarchical or “republican”) the final distribution of economic assets within society and across different societies. The third part turns to explore the effects of technological shocks in “modern” political economies, that is, those societies already endowed with

1 Aristotle, Politics, IV, p. 11; Machiavelli, Discourses on the First Ten Books of Titu Livy, Book I, Ch. 55.

state structures. Making use of the basic building blocks of the most recent literature on positive political economy, it suggests an analytical framework in which the “decisive” voter (who differs across political regimes) accepts or blocks a given technological shock as a function of its effect on her net income. The decisive voter’s response then determines the overall distribution of income. The model is employed to give a coherent account of some broad historical trends in the evolution of income inequality.

STATE OF THE LITERATURE

Accounting Models of Income Distribution

A first and quite extended approach to explain inequality consists in depicting any income distribution as a function of both the distribution of factor endowments among individuals and the prices paid to factors. As an accounting exercise, this research strategy has considerable empirical purchase: It generates valuable descriptions of different “moments” of inequality. As a result, it has been directly linked to the analysis of factor shares in national incomes. It has been employed also within the tradition of general equilibrium models to examine the impact of technological change on the distribution of human capital and therefore societal income, of population shifts from rural to urban labor markets, and of international trade on rewards to different economic factors. Atkinson & Bourguignon (2000, pp. 5–13) provide an excellent summary of this class of theories.

Factor-share models of income distribution run, however, against two theoretical and empirical roadblocks. First, they take as given the distribution of factor endowments among individuals at each period or slice of time, and therefore they can make sense of the evolution of inequality only in conjunction with some exogenous theory of economic change. This is unsatisfactory for those who wish to determine the extent to which inequality affects growth (either directly or through its impact on policies and institutions affecting development).

Second, factor-share models of income distribution assume mechanisms of price determination (mostly competitive markets) that are far too restrictive for the purposes of explaining income inequality. A key point of this article is that political and economic agents impose barriers to entry and hence distort markets to capture rents, which then lead to more or less inequality.

Take Kuznets’ path-breaking 1955 paper on economic growth and income inequality as an example of this line of inquiry. (Kuznets’ paper is in fact simpler than standard factor-share theories of income distribution because it does not even endogenize how a shift in factor sizes may affect the rate of return to each factor in equilibrium.) According to Kuznets (1955), the process of economic development triggers a population shift from rural to nonagricultural sectors. Given certain assumptions about the mean and variance of income in each sector, growth can lead to a temporary rise in overall inequality. This famous Kuznets conjecture about the curvilinear relationship between inequality and development has spurred an extraordinary amount of empirical research, albeit with rather inconclusive results. But it does not contain a complete theory of the emergence and dynamics of inequality—it simply relies on some kind of exogenous technological shock that generates changes in factor sizes and incomes.

Theories of (Intergenerational) Persistence of Inequality

The bulk of the contemporary economic research on inequality has been devoted to the construction of models to explain the temporal (intergenerational) persistence of inequality.

In their simplest version, these models take the following function form:

\[ k_{i,t+1} = f(k_{i,t}) - c_{i,t} + \varepsilon_t, \]

The lack of consistent empirical results is unsurprising. Kuznets himself stressed that development does not necessarily lead to more inequality in all instances. For a summary and discussion of existing evidence on the Kuznets curve, see Lindert (2000).
where \( k \) represents the assets of individual \( i \), \( c \) is the fraction any individual \( i \) consumes at period \( t \), and \( \varepsilon \) is some idiosyncratic shock that may affect assets \( k \) at time \( t \).

The most general (and plausible) interpretation of this functional form is a world where there is some intergenerational transmission of wealth among families or individuals. Individuals have some initial wealth \( k \), inherited from their parents, which they allocate to maximize their expected wealth at the end of the period. After consuming some of their wealth, they bequeath the rest to their children.

The main point of contention among researchers working with this class of models has to do with the particular functional form that Equation 1 takes—and, particularly, with the behavior of the function \( f(.) \) and the parameter \( \varepsilon \).

The parameter \( \varepsilon \) embodies what Becker & Tomes (1979) refer to as “inequality in luck.” It may be identical for all agents, or it may be specific to each one and correlated with some particular trait such as income. With particular values of \( f(.) \)—those that do not amplify inequality over time—\( \varepsilon \) has been shown to generate, in the long run, a unique limiting distribution (with perfect mobility and no poverty traps).

More important, the structure of the production function \( f(.) \) and of investment decisions determine the extent to which inequality declines, persists, or even rises over time. Several models portray the processes of production and investment as having an equalization effect on the distribution of income. The simplest way of generating this result is, following Solow’s aggregate neoclassical growth model, to assume a world in which each individual invests \( k \) and where there is a common production function \( f(k) \) that is strictly increasing and concave. With no credit markets in place, the expected marginal return on investment is higher for poor entrepreneurs, and the average wealth of poor families grows faster than the average wealth of rich families. Provided there are no shocks of the type embodied by \( \varepsilon \), the wealth of every family converges to a steady state with wealth \( \tilde{k} \). Inequality disappears in the long run.

(Naturally, these results depend on rather strict assumptions. If we assume that the propensity to save varies in line with income, the process of equalization does not happen in the long run.)

Under a different scenario, the distribution of income remains unchanged (in its variance) or even widens. Imagine, in line with the most common solution in the literature, that all or some investments leading to a high-return technology are by nature indivisible, and that any individual needs some \( k \) larger than some threshold \( k^* \) to acquire that high-return technology. Once more, assume credit markets are missing.\(^4\)

Such a world may then lead to an unchanged or even to a more polarized income distribution. Those with assets \( k < k^* \) are constrained to invest in low-return technologies and to remain in low income brackets. Those with \( k > k^* \) maintain or even amplify their advantage in the income distribution. Naturally, a few changes in the underlying assumptions can overturn the prediction of higher inequality. For example, those with few assets may save and accumulate enough wealth to overcome the threshold \( k^* \). The economy (and therefore average income) may grow over time to the point where all individuals end up with \( k > k^* \) and therefore can invest in high-return technologies.

Finally, technological shocks occurring with some probability \( \varepsilon \) may rearrange the positions of individuals in the initial income distribution.

These models overcome one of the central weaknesses of factor-share models of income distribution: They offer an endogenous theory of growth and inequality. Inequality, defined as the distribution of wealth above or below \( k^* \), becomes now a direct cause of economic stagnation. By contrast, more equal income distributions, and particularly distributions that make sure that most individuals have \( k > k^* \), foster the process of development.\(^5\)


\(^5\)This discussion draws very heavily from Kahhat’s (2007) excellent survey of the literature modeling the persistence of
Whereas factor-share models shy away from modeling the underlying causes of income inequality, the theories on the intergenerational transmission of wealth explicitly attempt to endogenize the evolution of inequality and its consequences for development, that is, for average income. They do so mainly to account for the growth effects of inequality—and only secondarily to address the causes of inequality per se. However, their effort has two shortcomings. First, they do not endogenize the initial distribution of income. Second, they do not offer a causal explanation of the functional structure (production function, distribution of idiosyncratic shocks, types of investment) dictating the generation and transmission of wealth. For example, almost all intergenerational models do not endogenize the presence of imperfect credit markets. Yet we know that institutional and legal frameworks vary across societies. Many economies may not have solved informational asymmetries that, under conditions of high income inequality, lead to further underinvestment and income polarization. But other economies have. To put it differently, financial-market imperfections (or, more generally, the presence of barriers to entry in markets) vary because they do not happen in a deus ex machina fashion. They are shaped by specific policy choices—often due to the particular distributional consequences of those choices.6

Institutional Models

In contrast to economic models of intergenerational transmission of wealth, where economic agents simply maximize income within a given market structure, several researchers have built a third class of models predicated on the notion that institutional and political choices matter. These models come in a variety of forms. Some stress the role of parties and policy choices to explain inequality (Hibbs 1977, Huber & Stephens 2001, Bartels 2008). An abundant literature on labor markets has related cross-national differences in the structure of wage bargaining with the level of wage compression—Wallerstein (1999) is likely to be the best exponent of this line of thought. Finally, the organization of welfare states has been linked to different poverty levels and income distributions (Esping-Andersen 1990, Goodin et al. 1999, Smeeding 2005).

Although these political-institutional models have made definite contributions to the debate on the sources of inequality, they have several shortcomings. First, their focus is limited to the evolution of advanced countries since the 1970s. Second, government partisanship and wage institutions have been found to have little effect on inequality in the long run, that is, once one looks at pre-1970 data (Scheve & Stasavage 2009). Finally, they are incomplete in the following sense: They do not integrate their theory of inequality determination within a model that explains the conditions under which a given distribution of income is compatible (or not) with the efficient allocation of resources and hence with growth.

ORIGINS OF INEQUALITY

This section inquires into the forces underlying the emergence of inequality, summarizing the full-fledged model and empirics presented by Boix (2010).

Primitive Societies and Self-Enforcing Peace

To understand what causes the emergence of inequality, let us start with a world populated by agents with the same resources (i.e., labor endowments and technologies of production). These agents can be thought of as individuals living together or as individuals or households
populating some territory. They can also be thought of as representative individuals of homogeneous groups, that is, groups formed by “identical” agents who have somehow solved their intragroup collective action problems. (As shown later, the possibility of several individuals or households living together is based on the same principles that make it possible for different groups to live in peace.)

Although those individuals still are, in line with standard economic assumptions, strict income maximizers, they may choose among two (stylized) alternative strategies to advance their goal. On the one hand, they may devote their resources to some direct “productive” activity such as hunting. On the other hand, they may employ them in predatory activities to grab other individuals’ production. In more complex societies, these two strategies should be thought of in broader terms. A productive strategy implies allocating one’s own time and endowments to produce goods and services, which are paid at the rate established in an arena of voluntary transactions—a market. An extractive or expropriatory strategy is equivalent to the appropriation of the assets or returns of other individuals, either directly or through the introduction of distortionary policies that change the prices of their assets. Extractive strategies take many forms: they may consist in regulating the supply of factors (through labor-market legislation, migration policies, capital controls, trade barriers to goods and services, etc.), setting prices at levels that differ from the ones set from purely voluntary transactions, or directly enforcing transfers from some agents to other individuals.

Assume that stealing is more profitable to one particular individual than producing if all the other agents are purely engaged in a productive activity. He may still profit from some of his own production while grabbing, without much effort or opposition, the output of the other player(s). By contrast, being a producer who is looted by the rest is the worst possible scenario for all agents. In between those two extreme cases, players prefer a situation in which everyone produces to one in which everyone loots (or attempts to loot) everyone else.

In this game, which has a prisoner’s dilemma structure, players are identical and their payoffs are symmetrical. As is well known, if individuals interact only once, generalized looting is the resulting equilibrium. However, if the game is played over time (with future payoffs valued with discount rate $\delta$), such as the ones represented in Figure 1, two equilibria may emerge. On the one hand, the players may

![Figure 1](image_url)

An iterated prisoner’s dilemma.
directly choose a looting strategy, which would then be self-sustaining over time since every player has an incentive to punish the other player in response to his noncooperative response. On the other hand, they may follow a trigger strategy according to which they choose a productive response to start with and only shift to looting if the other agent loots them.

The latter strategy leads to a sustainable production-production equilibrium if the value of choosing a production strategy is larger than that of looting a producer in the first period and then facing a looting solution ever after. Employing the payoffs in Figure 1, it is an equilibrium if $3/(1-\delta) > 4 + 2(\delta/1-\delta)$. Solving this inequality shows us that peace and production are possible whenever $\delta > 1/2$, that is, whenever the players do not discount the future too heavily.

There are two key lessons in this game. The first one, which is well known in the literature, is that even in a Hobbesian world, where individuals have strong incentives to raid other agents, it is possible for all players to avoid conflict, sustain peace, and engage in productive activities without having to resort to any centralized mechanisms of authority. Anarchy and peace are compatible because everyone has an incentive to behave well in response to everyone else also behaving well.

The second lesson of the game, which is central to this article's analysis, is that a self-enforcing peace outcome can take place only if there is some fundamental equality of conditions among players. To see why, let us examine what would happen if individual $B$ derived a higher payoff from producing (6 instead of 3) as a result of experiencing a positive technological shock. Assume moreover that all of the increase would be absorbed by the looter; that is, that if $A$ decided to raid $B$, $A$'s payoff from looting would grow from, say, 4 to 7. In this new context, $A$ would follow a peaceful strategy only if $3/(1-\delta) > 7 + 2(\delta/1-\delta)$. Given this inequality, $A$ would choose a production strategy only if $\delta > 4/5$, that is, if it values future payoffs very highly indeed. (By contrast, $B$'s incentive to follow a production strategy increases as the production payoff goes up. In this particular example, for any $\delta > 1/5$, that is, even for cases in which the future is heavily discounted, $B$ prefers peace.) In other words, as inequality increases, the incentive to cooperate declines for the less advantaged side, and systematic conflict becomes much more likely.

The main insights of the game match well the functioning characteristics of primitive, preagrarian communities. According to archaeological evidence and anthropological research, foraging communities are small—the sum of a few families. Their internal level of differentiation is minimal. Individuals exhibit similar age-specific patterns of consumption. Equality persists over time owing to the type of technologies in use (strongly correlated with individual ingenuity and physical strength) and to the fact that asset accumulation (and hence any intergenerational transmission of wealth) is impossible. Cooperation takes place without stable structures of authority or permanent leaders, at least beyond one generation. Band or tribe chiefs act as mere referees among different individuals or families, cajoling, persuading, or mediating between their group members (Clastres 1972, 1974). As emphasized above, non-institutionalized peace, that is, peace unenforced by any third party, is only one of the possible equilibria of the game. Even under conditions of equality, a fleeting shift in discount rates or in the technology of looting may result in the collapse of cooperation. In fact, we know that generalized violence is endemic among preagrarian societies (Gat 2000, LeBlanc 2003). But foraging bands manage interpersonal conflict through a regular process of fission, with a section of the group seceding and migrating to a new location, where it self-organizes again along the standard patterns of relative equality and non-institutionalized politics (Hirschman 1981).

### Technological Shocks, Emergence of Inequality, and Political Violence

The initial equality of conditions, and hence the possibility of self-enforced peace, breaks
down as soon as there is some kind of technological shock—primarily the domestication of plants and animals—that raises the payoffs differently across different individuals, households, or groups.

The immediate sources of that technological shock, that is, of the agrarian revolution, may have been manifold and should not concern us here. It may have been driven by particular environmental changes; it may have been endogenous to population growth, which may have resulted in a downward shift in the marginal productivity of hunting and gathering; it may have been the outcome of a slow process of learning-by-doing; or it may have been random and unexpected (Watson 1995, Bar-Yosef & Meadow 1995). Nonetheless, as emphasized by Diamond (1997) and tested by Hibbs & Olsson (2004), independently of all these causes, the adoption of agriculture only occurred in those particular regions of Earth that met some particular biological and geographical conditions, such as an appropriate climate and latitude.

The agrarian revolution triggered a momentous transition. Formerly, all men were engaged in foraging activities, and the marginal productivity of hunting was the same across the board. The new technologies of storage and plant domestication led to a high degree of territorial variation in terms of land fertility and productivity. In other words, the introduction of new technologies caused an initially homogeneous world to acquire a differentiated resource gradient with a core of rich lands and a periphery of marginal quality. Those who happened to occupy the core now obtained a much larger income than the rest.

As already discussed, the technology-biased shock and the correlated increase in inequality has a discriminating effect on the behavior of individuals. Whereas the agent who benefits from the shock still has an incentive to follow a production strategy, the agent who does not has a much higher propensity to loot. In other words, as inequality increases, violence becomes more likely both within established communities (those sharing a given territory) and across territorially differentiated communities. Moreover, violence takes place in a particular direction: from the disadvantaged to the advantaged. (Naturally, violence does not only happen for those motives and in that direction. For example, the “natural producers” may have an incentive to respond in kind to protect themselves.) This violence may occur either within human groups that share a compact territory, or between groups. The latter case matches a rather strong regularity in human history: the fact that peripheral, less rich lands have tended to breed much more warprone societies than fertile lands. Ancient Mesopotamia and Egypt, the first centers of agriculture in the Middle East, withstood successive waves of invaders originating in the Asian steppes. The Roman Empire fell under the pressure of Germanic tribes. Medieval Europe endured the attacks of periphery populations such as the Vikings, Slavs, Turks, and Mongols.

**State Formation and Political Sources of Inequality**

The rise of economic inequality and violence induced, in turn, a key transformation in the existing political structure, which then affected the distribution of wealth and income.

**Institutional Solutions to the Problem of Violence.** To sustain peace and the possibility of a fully productive strategy (where joint gains are certainly maximized), individuals have two alternatives. Both of them involve the formation of some organization or structure that has the monopoly of violence over a given territory—that is, they involve the construction of a “state.” But the institutional mechanisms that underpin each solution are very different.

On the one hand, the producers—a term that I use now to define those who benefit the most from the technological shock—can transfer part of their production to the potential looters or bandits (those who do not benefit from the shock), in exchange for which the latter refrain from violence and supply some protection against other bandits. (Bandits and looters need not be external enemies. These terms apply as
well to residents of the community who may try to steal from, free-ride on, or otherwise take advantage of richer individuals. This first path to peace, which we may refer to as a “monarchical solution,” follows from Olson’s (1993, 2000) insight that the state emerges as “roving bandits” turn into “stationary bandits” or landlords.\footnote{Writing from an anthropological perspective, Carneiro (1970) also offered a theory of state formation that emphasized the interaction of violence and exit options (the latter deriving from population density). Exit strategies and density were, however, not endogenized in his paper. Like Olson, he focused on the emergence of autocratic states. See Olson (1977) for a review of state-formation theories in anthropology and archaeology.}

On the other hand, the producers may simply spend some part of their own resources to set up institutions or structures, such as a common army and stable leadership, to defend their assets and lives against bandits. This second path to state formation, where producers also double as defenders and rulers, implies, in a word, the construction of a “self-governing” polity. It is important to stress here that a system in which the community of producers sets aside some portion of its income to hire an army or leader to protect its members (while governing themselves through “republican” institutions) does not constitute a case of self-government or republicanism. Since the hired leader cannot credibly commit to preserve the terms of the contract once he has been appointed and takes control of the army, hiring a protector or condottiero ends in (is equivalent to) the subjection of the producers to a leader-monarch. In other words, a stable republican solution is one in which citizens have ultimate control over the means of defense.

This self-governing strategy, based on spending some resources to deter looters, comes in two variants. On the one hand, the producers may mainly focus on production and then spend a minimal portion $\theta_D$ of their income for purely defensive purposes, such as building a wall or a watch tower to observe the horizon. On the other hand, the producers may decide to increase the amount of resources they spend on military endeavors, beyond what is needed to defend their territory, to attack the looters. The offensive strategy costs an extra portion $\theta_O$ of their income and puts overall spending at $\theta = \theta_D + \theta_O$.\footnote{It seems plausible that even when the main endeavor of these agents is production, they still need to pay some minimum costs of defense to be ready to move to the fully militarized strategy. If the latter were equal to 0, then government institutions would appear only when bandits actually appeared on the horizon. When they were not there, everyone would simply cooperate spontaneously with no permanent structure governing them. This does not seem realistic. For a formal proof of the condition under which this solution prevails, see Boix (2010).}

If the offensive costs $\theta_O$ are high, the producers will be more likely to follow a strictly defensive strategy. However, as they converge to 0, the value of looting outsiders becomes higher than the costs of attacking, and the self-governing community becomes an imperial republic in both its nature and its behavior.

**The Distributive Effects of Different Political Regimes.** Each political solution comes with very different distributional consequences. Let us work out the main logic of the argument using the payoff structure discussed in the subsection “Primitive Societies and Self-Enforcing Peace.” Before the state was established, if peace was an equilibrium, each individual received a production payoff of 3 and the distribution of income was equal. Now, after a technologically biased shock takes place, the distribution of income becomes unequal: Individuals $A$ make twice as much as individuals $B$. However, this new income distribution is not the final one because it triggers a set of political changes that reshape the material position of individuals. In a monarchical solution, predicated on extracting a portion $\epsilon$ of the assets or income of producer $A$ and siphoning it to the potential bandit $B$ such that the latter prefers not to pillage $A$ (and protects $A$), each natural producer keeps $(1-\epsilon)6$, and the bandit-turned-ruler receives $3+6\epsilon N$ where $N$ is the number of strict producers.\footnote{The parameter $\epsilon$ should be such that both the ruler and the producers are better off than in a looting-looting equilibrium. Moreover, the monarchical settlement is only feasible if peace} By contrast, in a republican or
Table 1 | A simulation of the effects of monarchy

<table>
<thead>
<tr>
<th>Table 1 A simulation of the effects of monarchy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Baseline</td>
</tr>
<tr>
<td>Monarchy</td>
</tr>
<tr>
<td>Low extr.</td>
</tr>
<tr>
<td>High extr.</td>
</tr>
</tbody>
</table>

Table 1 simulates the effects of establishing a monarchical world and a self-governing regime once the initial, equal world experiences a technological shock. Individuals of type A constitute 90% of the population and double their output from 3 to 6. For a given discount rate $\delta = 0.5$, the shock makes the spontaneous peace equilibrium unfeasible. Assume that individuals of type B have a better military technology than type A such that they prevail and impose a monarchy. Given the structure of the model, the extraction rate $\varepsilon$ must range between 0.25 and 0.50 for the monarchical settlement to be an equilibrium for both parties. I use this range to calculate the postextraction per capita income of individuals A and B, the proportion of total income in the hands of the top 10%, and the Gini index.

Table 1 shows that monarchies have a strong effect on the distribution of income. The ratio between rulers and producers goes from 0.5 after the shock but before the establishment of a state to 3.7 under the lowest extraction rate ($\varepsilon = 0.25$) and 10 under the highest extraction rate ($\varepsilon = 0.50$). The Gini index goes up from 0.05 after the shock (yet before the introduction of a state) to 0.18 if the extraction rate is low and 0.42 if the latter is high. Although not shown here, holding the feasible rate of extraction constant, the smaller the number of rulers relative to the population of the country, the more unequal the distribution becomes. For example, halving the ruling group to just 5% of the total population doubles the rulers/producers income ratio.

The distribution of income will be also shaped by the way in which rulers extract their resources or, in other words, by the kinds of alliances they form with different strata of producers. The distribution of income becomes highly skewed when the rulers ally with a particular fraction of the population (imposing a lower extraction rate $\varepsilon$ on them than on other groups) or when the ruling elite simply replace the existing segment of wealthy individuals. By contrast, the ruling elite may favor a more equalized structure of income within the producers’ population. These different outcomes, resulting from different institutional and distributive choices, are particularly manifest in processes of repopulation and colonization. After the Christian kingdoms of northern Spain took over most of the southern half of the Iberian Peninsula in the thirteenth century, they differed markedly in their strategies of land state. Second, it will vary with the tax elasticity of output and with asset specificity (Boix 2003). The higher the mobility of assets, the less punitive the level of extraction by the lord. Hence, predation and social and economic inequalities will be sharper in agrarian economies than in urban settings (for identical distributions of technologies across individuals).
redistribution: Castile and Portugal favored the formation of large landholdings; Catalans settled in Valencia as small farmers (Vicens Vives 1957). Similarly, the conquest of the Americas followed very divergent paths: Farming communities settled in the northeastern seaboard and in Quebec; the Caribbean basin and Central and South America were dominated by landowners (Engerman & Sokoloff 2002). In all those cases, the distribution of land partly responded to economic considerations, such as the supply of labor. But it was strongly affected by the way in which the ruling elites decided to structure the distribution of assets—and that was probably related to the prevailing political institutions and societal conditions back in the conquering countries.

Under a nonmonarchical system, final equality within the producing, self-governing community will closely mirror the technologically induced inequality (provided taxation to pay for defense is not progressive): The overall distribution, including the looters, would be actually lower because the producers have to spend some of their resources on deterrence. Although a monarchical solution has a much stronger effect on the distribution of income resulting from the technological shock than the republican solution, this does not mean that monarchical regimes should always have more skewed income distributions than republican governments. Consider a case in which the technologically induced distribution is quite unequal but still within a range that makes cooperation among individuals a feasible outcome. If the extraction rate under a monarchical regime is sufficiently low, then it may well be that a republic regime is more unequal than a monarchical structure, particularly given that in a monarchical solution resources are transferred from technology-rich to technology-poor individuals.

**Some Empirical Evidence.** From an empirical point of view, the simulated extraction rates and Gini indexes do not seem farfetched. Consider the evidence gathered by Milanovic et al. (2007) and reproduced in Table 2. In almost all preindustrial regimes, the Gini index fluctuates around 0.40, and in most cases the top 10% of the distribution controls more than two fifths of total income. Bourguignon & Morrisson (2002) offer similar data across the world in 1820, at the dawn of the industrial revolution.

Owing to the lack of reliable data on income distribution before the nineteenth century, and given that height and individual income seem to be correlated (at least within ethnically similar groups) (Steckel 1995), Boix & Rosenbluth (2006) employ height as a proxy for access to food resources, health status, and general welfare. In preindustrial economies, insufficient nutrition was widespread, mostly affecting the poorest social strata. Whereas the top decile consumed 4329 kcal per day in England in 1790, the bottom decile consumed 1545 kcal per day (Fogel 1994). Hence the distribution of heights should provide us with clues about how that society distributed its resources, at least in human communities living under conditions of scarcity.

As summarized in Table 3, inequality is low within tribal, preagrarian groups. With the transition to a fully agricultural society with complex political institutions, heights vary strongly with the type of political regime. In ancient kingdoms and absolute monarchies, there is considerable differentiation among individuals: The height gap between nobles and peasants ranges from 7 to 9 cm. Using fitted models between log income (in constant dollars of 1985) and height (Steckel 1995), this difference translates into an income ratio of 5 or 6 to

<table>
<thead>
<tr>
<th>Society</th>
<th>Gini index</th>
<th>Top 10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roman Empire ca. 14 AD</td>
<td>0.36</td>
<td>0.44</td>
</tr>
<tr>
<td>Byzantium 1000</td>
<td>0.41</td>
<td>0.46</td>
</tr>
<tr>
<td>England/Wales 1688</td>
<td>0.45</td>
<td>0.38</td>
</tr>
<tr>
<td>Moghul India 1750</td>
<td>0.39</td>
<td>0.35</td>
</tr>
<tr>
<td>Mexico 1790</td>
<td>0.64</td>
<td>0.61</td>
</tr>
<tr>
<td>Naples 1811</td>
<td>0.28</td>
<td>0.35</td>
</tr>
<tr>
<td>Brazil 1872</td>
<td>0.39</td>
<td>–</td>
</tr>
<tr>
<td>France 1780</td>
<td>–</td>
<td>0.56</td>
</tr>
</tbody>
</table>

Sources: *Milanovic et al. (2007); **Morisson 2000.
Table 3  Political settings and differences in male heights

<table>
<thead>
<tr>
<th>Institutions and places</th>
<th>Mean (in cm)</th>
<th>Difference (in cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-State Society</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zuni Pueblos (before 1680 AD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height at 90th centile</td>
<td>165.8</td>
<td>4.7</td>
</tr>
<tr>
<td>Height at 50th centile</td>
<td>161.1</td>
<td></td>
</tr>
<tr>
<td><strong>Monarchical/Authoritarian Societies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ancient Egypt (New Kingdom)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Royal</td>
<td>174</td>
<td>8.0</td>
</tr>
<tr>
<td>Commoners</td>
<td>166</td>
<td></td>
</tr>
<tr>
<td>Mycenae</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Royal</td>
<td>172.5</td>
<td>6.4</td>
</tr>
<tr>
<td>Commoners</td>
<td>166.1</td>
<td></td>
</tr>
<tr>
<td>Poland—17th century</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noble</td>
<td>170.6</td>
<td>7.7</td>
</tr>
<tr>
<td>Rural non-Jewish</td>
<td>165.2</td>
<td></td>
</tr>
<tr>
<td>Jewish</td>
<td>162.9</td>
<td></td>
</tr>
<tr>
<td><strong>Democratic Society</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Members of Ohio National Guard, mid 19th century</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laborers</td>
<td>173.3</td>
<td>2.2</td>
</tr>
<tr>
<td>Skilled workers</td>
<td>174.0</td>
<td></td>
</tr>
<tr>
<td>Farmers</td>
<td>174.7</td>
<td></td>
</tr>
<tr>
<td>Professionals</td>
<td>175.5</td>
<td></td>
</tr>
</tbody>
</table>


1 between the two groups. By contrast, more democratic settings, such as preindustrial Ohio, show a much more compressed height structure.

**Endogenizing State Formation**

Although a detailed discussion of the factors that ultimately determine the choice of political regime is beyond the scope of this article, which regime will prevail is related to the costs it imposes on producers. Producers will opt for a self-governing community if $\epsilon > 0$. Otherwise—that is, if the costs of defense are higher than the taxes they pay—the community of producers will end up subject to the bandit-lord.

Self-governing institutions will be more likely when the exercise of violence and war does not rely on sophisticated weaponry (such as swords, chariots, horses and so on), which promotes specialization and leads to the formation of a separate caste of warriors, and when producers live in territories with geographical conditions (such as sea barriers and mountainous terrains) that make it easier for producers to deter bandits.

As weapons become more complex and expensive and as organization (in large armies) becomes central to military action, the production of violence will turn into a rather specialized activity, giving an important advantage to those individuals who decide to engage in predatory activities. Strong and centralized states, highly extractive fiscal systems, and rather unequal societies will then prevail.

The foundation of the first big states in Mesopotamia around 3500 BC was probably related to the introduction of bronze weapons. The introduction of two-wheeled chariots around 1800 BC increased the costs of war and led to the formation of a set of “feudal” states with a narrow aristocracy of warriors exercising full control over their subjects (McNeill 1982. As emphasized by Rogowski & McRae (2008), medieval feudalism followed the introduction of the stirrup and the strengthening of the cavalry, and absolutism was equally related to the sixteenth-century revolution associated with the use of cannons and standing armies.

Military technology has not always evolved to trigger more centralized and hierarchical structures. Occasionally, there have been organizational and mechanical changes that have reduced existing imbalances between warriors and producers. Such changes have resulted in a broadening of the political base of government (to the point of “democratizing” institutions) and the re-equalization of the distribution of income. In the second half of the second millennium before Christ, the discovery of how to make weapons out of iron cheapened war, allowed a “relatively large proportion of the male population [to] acquire metal arms and armor” (McNeill 1982, p. 12), and gave way to more egalitarian political and social structures. Aristotle wrote that “when the country is adapted for cavalry, then a strong oligarchy
is likely to be established,” whereas “the light-armed and the naval element are wholly democratic” (Politics, Book 6, Part 7). Levi (1997) emphasizes the impact of modern war and general military mobilization on the extension of political rights, and several authors have linked the occurrence of modern war to higher levels of taxation and redistribution (Peacock & Wiseman 1961).

PERSISTENCE OF INEQUALITY
After the formation of the state, the overall distribution of income continues to be a function of both technological shocks (changing the supply and demand of factors and their ownership and prices) and military and political factors (such as the domestic balance of power, the internal organization of the state, and the nature of interstate competition).

Nonetheless, there is a key difference between stateless communities and societies governed by states. In the latter, technological shocks affect the distribution of income embedded in a stable, institutionalized political structure. That is, they are mediated by political procedures according to which a relatively stable set of individuals systematically vote over policies and have enough authority to implement them. In primitive or prepolitical societies, technological change happens in an unconstrained manner—and often ends up wiped out by conflict unless some institutions are constructed to protect its application and gains. In a political society, the decisive voter (a dictator, a narrow ruling elite, or the electorate of a democracy) chooses the particular policy or regulatory framework that determines (even if partially) access to new technologies and, therefore, the productivity and income of all economic agents.

The choice of a particular regulatory or policy framework is equivalent to the imposition (or elimination) of entry barriers to the technologies of production. The concept of policy regulation should be understood here in a broad sense: It may imply the introduction (or abolition) of high tariffs, the passage of particular laws that bar the use of new transportation technologies, the monopolistic (or competitive) regulation of a given market, and so on. As stressed above, ancien régime states tended to impose those barriers in a discriminatory fashion precisely to glue together the social and political coalitions that sustained them.

In a politically institutionalized society, the decisive voter (what standard spatial models identify as the “median voter”) compares her income under different technological alternatives and chooses the regulatory framework that maximizes her income. She will seek to block policy reforms (such as the suppression of guilds, the abolition of capital controls, or trade liberalization) that would reduce her final disposable income and promote those that would increase it.

Notice that the choice of the policy parameter will not be simply based on the “direct” effect it has on the income of the decisive voter (or voters). A given regulatory framework shapes the pretax income of all individuals, that is, it generates a particular income distribution. If we accept existing optimal taxation models (Meltzer & Richards 1981) and their application to the choice of political regimes (Boix 2003, Boix 2008), each income distribution will be associated with a tax rate, and, given some costs of excluding part of the population from participating, with a particular political regime. Hence, in choosing a particular policy framework, the decisive voter calculates the change in her final income, taking into account not only the direct economic effect (a change in the production function) but also the taxes she will have to pay (under each political regime) and the costs of excluding part of the population (if any) to retain her position as a decisive voter.

In the following two subsections, I examine the choices of the decisive voter. I employ a simple model in which, in the spirit of policy-reform models such as Fernández & Rodrik’s (1991), there are two types of individuals, $b$ and $l$, with their respective incomes $y_b > y_l$, and the $l$ individuals constitute a majority. In the first subsection (“Political Reform and Inequality in Authoritarian Regimes”), I consider the
reaction of high-income individuals to biased technological shocks when they are the decisive voters; because they are a minority, such a case corresponds to an authoritarian regime. In the next subsection (“Political Reform and Inequality in a Fully Enfranchised Polity”), I explore the responses to similar shocks when the decisive voters are the low-income voters. I then add some complexity to this exploration (“Amendments”) by considering three extensions: policy making under uncertainty, the choice of a policy framework in the context of multiple countries, and the effect of relative gains.

One of the findings of the following discussion is that underdevelopment is caused by the reluctance of preindustrial political elites to introduce any policy reforms. Similar results have been derived by Olson (1982), Justman & Gradstein (1999), Krusell & R´ıos-Rull (1996), and Parente & Prescott (1999), who insist that the loss of economic rents leads elites to block reforms, and Acemoglu & Robinson (2001), who have added elites’ concern over the loss of political power to explain the failure of economic liberalization. The approach I suggest here allows us to explain also why inequality is part and parcel of that suboptimal outcome of authoritarianism and underdevelopment, tease out the political and economic sources of stagnation, and explore the particular political environment that facilitated the industrial breakthrough. Moreover, it offers a way to treat in a unified structure both the instances of policy blockage inherent to many ancien r´egime elites and the opposition to change in many democratic societies.

Political Reform and Inequality in Authoritarian Regimes

Political reform and inequality under potentially falling incomes. Consider first a case in which only high-income voters decide. To exclude the rest of the population, they need to pay some exclusionary or repression cost $c$—but since this cost is lower than the tax they would pay if everybody voted, they impose an authoritarian regime. (The cost $c$ is lower because the level of inequality is high enough to lead to higher taxes under democracy. Throughout the discussion I assume $c$ increases with both the size and the income of the excluded population. Poor populations are easier to repress the poorer they are. For a formal analysis, see Boix 2003.)

The decisive voter, a high-income individual, will block any reform that leads to a fall of his pretax income—unless the regulatory change leads to a distribution and type of wealth such that a tax under democracy is lower than the costs of excluding low-income voters under the status quo. An example of that potential shock would be the introduction of new commercial practices or inventions in a strictly agrarian economy that reduce the value of existing assets (such as land) or that divert labor to new sectors (and therefore lower labor supply).

A very stylized representation of the general case of maintaining the status quo appears in Figure 2a. If high-income individuals pass the reform, they suffer an income loss. This loss is not compensated by a possible change to a broader franchise (as a result of more equal conditions and less threatening taxes) that would reduce $c$ to 0. Hence, no change is implemented.

As I point out later, this instance of blocked reform was pervasive among ancien r´egime systems. In preindustrialized societies, wealthy individuals owned very specific assets, that is, assets whose value hinged on regulatory barriers sustained by the state and/or wealth (such as land) that could not be easily redeployed in new, more productive economic sectors. In fact, their position on the social ladder derived precisely from the types of highly extractive institutions that accompanied the formation of the state. Because inequality and authoritarianism came hand in hand, any economic change threatened the political basis of the wealth of the ruling elite. Hence, those old regimes were logically associated with long-run stagnation.

Political Reform and Inequality under Potentially Increasing Incomes. By contrast, the decisive voter may support a reform or accept a shock if the change in regulation or
the economic transformation boosts his pretax income. Still, his final decision will depend on the overall effects of the potential reform on the whole population.

Consider first the case in which income grows only among high-income voters. Given that the exclusion costs \( c \) do not change (by our assumption that they are determined by the income level of \( i \) individuals), the decisive voter accepts the policy change while maintaining the exclusionary political system. The discovery of natural resources (such as oil) or commodity price booms fit this case: They increase inequality and reinforce authoritarianism, provided that those economic resources are owned by those in control of (or in alliance with) the state. [On the general effect of oil on political regimes, see Ross (2001) and Boix (2003); on the impact of price cycles, see Friedman (2006).]

Let us examine now the case of a technological or policy shock that increases incomes across the board. In those instances, the exclusion costs shift upward (given that low-income individuals have more resources available to them to oppose their exclusion from the ballot booth). Assume that exclusion costs go up linearly with the poor’s income, perhaps above some subsistence wage under which low-income
voters have no strength to oppose exclusionary measures. Then, under any income increase that does not change the overall level of inequality (that is, whenever both types get the same raise in relative terms), the decisive voter will approve the reform and maintain the political status quo of restrictive franchise, since he continues to have enough extra resources to contain the disenfranchised.

However, as soon as the income shock becomes “poor-biased,” that is, as soon as it increases incomes among low-income individuals proportionally more than among high-income voters, income inequality declines, lowering the cost of taxes (under full enfranchisement) relative to the cost of repression (while, as before, exclusion costs rise). If the poor-biased reform or shock is small, the decisive voter will decide to block the reform because, although it would lead to higher exclusionary costs, which would push high-income voters to prefer democracy (after the economic reform had been passed), the after-tax after-transfer income in that democratic setting would be probably lower than the net income under an authoritarian regime before the reform was passed (since inequality would still remain high and would result in considerable taxes). If the poor-biased shock is, instead, sufficiently strong, the postreform or postshock level of inequality will become moderate or low. That decline in inequality (jointly with rising exclusionary costs) will lead high-income voters to prefer paying taxes under democracy than funding an authoritarian police. However, that type of highly poor-biased shock—leading to growth, equality, and democracy—will be highly unlikely in closed, authoritarian political economies, particularly if the initial level of inequality is very high. Given these considerations, it seems reasonable that change could only come through some revolutionary shock, some miscalculated policy move by the governing elites, or a reaction to some external threat.

To summarize the discussion, permanent growth and considerable reductions in inequality, followed by peaceful transitions to democracy, are extremely elusive in ancien régime political systems. Because authoritarian regimes come into place by imposing unequal distributions of wealth and strong entry barriers to favor the ruling coalition and its allies, they are remarkably robust political equilibria. Statistically they have been: Democratic or republican polities account for less than 1% of all the states that emerged after the domestication of plants and animals, and it took more than 10,000 years to start the industrial revolution and to get democracy and relative levels of equality in a few parts of the world.

Political Reform and Inequality in a Fully Enfranchised Polity

In a fully enfranchised polity—one where low-income individual \( l \) is the decisive voter—the interests of the decisive voter are very often a mirror image of the interests of a high-income voter \( h \) in the authoritarian setting we just discussed.

Increasing Incomes for Low-Income Individuals. Any poor-biased shock or reform, i.e., any shock or reform that increases \( l \)’s income more than \( h \)’s income (Figure 3a), will receive the support of the decisive voter—unless the reform leads to a fall in taxes and transfers (resulting from shrinking inequality) that offsets the positive direct impact of the reform on his income.\(^{11}\) A “rich-biased” shock, one that boosts the income of high-income individuals more than poor, should not necessarily be opposed by \( l \): Low-income individuals may still experience direct income growth (and some gains in after-transfer income through the tax-and-transfer system). However, they will block a rich-biased shock if the latter generates enough inequality to turn authoritarianism into a feasible option for high-income voters.

Declining Incomes for Low-Income Individuals. Negative shocks, such as those “skill-biased” technological shocks experienced in

\(^{11}\) This case (a decline in transfers bigger than an increase in direct income) becomes more likely as high incomes fall in absolute terms.
industrial economies in the past 30 years, will be blocked by voters unless the direct fall in income is (automatically) compensated by higher transfers (Figure 3b). Since the introduction of full or quasi-full enfranchisement in the past 100 years, unskilled and semiskilled voters have delayed or impeded shocks and reforms that would hurt their welfare at several critical junctures. Examples include restrictive immigration policies and probably trade protectionism among Western nations in the 1920s and 1930s, the emergence of populism and ISI in Latin America, and the maintenance or implementation of strong labor regulations (leading to the formation of an insider-outsider split in labor markets) in many European countries since the 1970s. Antiliberal or anticapitalist revolts were common for example among guilds and workers in city-states of past ages.

The only way to pass proreform, liberalizing strategies under those circumstances (losses borne by the decisive voter) entails establishing some transfer scheme or compensatory package from the winners to the losers. This seems to account for the relatively well-established correlation between trade openness and larger public sectors (Katzenstein 1985, Rodrik 1998, Adserà & Boix 2002).
The likely time inconsistency of transfer schemes may explain why compensatory policy, and hence any reform that delivers or keeps low levels of inequality, happens mostly under institutions that strengthen the representation of affected voters, such as encompassing unions or proportional-representation rules (Rogowski & Kayser 2002). However, contrary to a widespread literature, not all compensation packages work optimally—only those that change the marginal productivity of low-income voters do (Boix 1998).

**Amendments**

**Uncertainty.** So far we have assumed a setup in which individuals have perfect information about the consequences (in terms of direct income, political regime, taxation, and transfers) of any alternative regulatory structure. Once we introduce the more realistic idea of an uncertain environment, the status quo becomes even more likely to prevail over time (Fernández & Rodrik 1991).

To see the stabilizing nature of imperfect information, consider first the case in which, although the decisive voter stands to gain from the reform, aggregate losses are larger than aggregate gains. If the decisive voter knows only about aggregate results (not about her individual outcome), in expectation she has losses, and therefore she blocks that reform. Examine now the reverse case—with total gains larger than total losses but with net individual losses for the decisive voter. Having gains in expectation, the decisive voter passes the reform. But after the reform is implemented and the decisive voter realizes her actual losses, she will repeal the reform—provided she still retains her pivotal position. The asymmetry between ex ante expected losses (leading to no reform) and ex post real losses (leading to reform reversal) reinforces the stability of the status quo.

**Interstate competition.** Let us now abandon our Robinson Crusoe model of politics in which our decisive voter chooses a regulatory framework for himself and his Friday without any concern for the decisions, growth rates, and income distributions of foreign powers. (Of course, we can also call it the Friday model of politics, where Friday chooses policy for himself and his master.) Consider instead what happens when our decisive voter has to make decisions surrounded by other countries.

Interstate competition spurs technological change. Assume that the decisive voter holds the belief that the probability of other countries successfully attacking his country (or simply intervening in its domestic affairs) increases with the ratio of total income of the foreign country over total domestic income. To pre-empt foreign attacks, countries may push for reforms even when they hurt the decisive voter (provided they deliver a larger total income). That they actually do push for such reforms, however, is not very likely. In a context of uncertainty about how those shocks or reforms may work, no state is likely to move unilaterally into uncharted waters. Still, the probability of reforms (and therefore of shifts in the distribution of wealth) will be higher the more fragmented the state system is. Hence, it was by virtue of its fragmentation that Europe had a competitive edge over China in the modern era (Jones 1981).

Provided no economy introduces a growth-enhancing policy reform, interstate competition does probably little to alter the status quo. However, as soon as one country takes the technological, economic, and therefore military lead over the rest, it plays a key role in the

---

12. No consideration has been given to growth either, or rather, growth is already subsumed in the model with individuals having zero discount rates—they only care about total income in their lifespan.

13. Or, following Fernández & Rodrik (1991) more strictly, aggregate losses are larger than aggregate gains for the economic sector (and not necessarily the whole economy) to which the decisive voter belongs.

14. This uncertainty about the future may explain the war bias of ancien régime states: They prefer to attack other states (something they know well how to do, since it is at the basis of the formation of an authoritarian polity) than to invest in generating autonomous sources of growth.
process of economic transformation of everyone. After England’s take-off became apparent after the 1830s, other nations scrambled to implement the liberal blueprint across Europe. As they did, they faced considerable internal conflict. The conservative-liberal divide that spread across Europe (like the one in the Middle East today) closely reflected a split between pro-status quo and pro-reform policy makers. But, overall, reform happened and in fact accelerated in response to a growing performance gap. Direct military threats or defeats made the income gap even more glaring and pushed some states to embrace reform quite abruptly, e.g., Meiji Japan. Countries that were unable or unwilling to respond either disappeared, engulfed by ever-growing colonial empires, or simply became marginal islands in the international economy.

Envy. So far, we have assumed the decisive voter chooses according to a strict principle of (after-tax, after-transfer) income maximization. But voters may decide with an interest in preserving (and increasing) their position vis-à-vis other individuals. The introduction of envy favors the status quo in polities dominated by high-income voters. They will oppose any shock that is not strictly rich-biased, that is, any policy that does not preserve or increase the existing level of inequality. Similarly, low-income voters will fight any shock that is not strictly poor-biased—even those that would lead to growing incomes in absolute terms. Hence, political economies (such as caste societies) or professions (such as academics?) mostly based on status should be particularly resistant to change.15

Contemporary Historical Trends in Inequality

Employing the analytical tools of the previous subsections, let us now impose some analytical structure to the evolution of income inequality, mostly in the past two centuries. (I focus on this period for two reasons. First, we have relatively reliable data for at least a few countries. Second, the level of income inequality has gone through substantial changes since 1800 relative to previous periods.)

In the long run, prosperity has bred equality. Figure 4 presents cross-sectional data of per capita incomes and Gini coefficients across the world after 1950. At low levels of development there is considerable variance in the overall distribution of income. At high levels of development variance is much smaller. Figures 5a and 5b show, in turn, the historical evolution of the income share of the top decile in several English-speaking and continental European countries for the past two to three centuries. Those countries with considerable inequality at an initial stage have undergone a substantial process of equalization over time. Figure 6 reports data on the wealth share of the top 1% in Switzerland, the United States, and the United Kingdom.

Industrial (and even commercial) capitalism, and therefore the breakdown of the agrarian economy, came hand in hand with a shift in the overall institutional or regulatory structure—from the statism and mercantilism of the ancien régime to a liberal economic order. Because the triggering event of modern economic development, the English industrial revolution, was a unique event (it was the first industrialization process, unconditioned by any previous industrial experiment), any explanatory theory may never be able to pass a standard test of significance—against the null hypothesis of pure chance. But although the causes of the industrial transformation are still heavily debated, two key factors, discussed in the previous subsections, contributed to it: the fragmented nature of the European state system, which fostered among policy makers a search for a successful strategy to grow faster than other powers; and, in that general context, the revolutionary break of 1688 in England that dislodged the existing reactionary or blocking coalition from power. Those two factors are

15Fehr et al. (2008) offer strong experimental evidence of envy in caste settings.

Inequality went up with the first phase of industrialization in Britain. According to the model sketched above, this should have actually helped to foster the process of economic transformation from a political point of view: the ruling elites were interested in those changes because they benefited them directly. To put it differently, this model turns the current explanation of the Kuznets curve on its head. In the standard account of the Kuznets conjecture, inequality resulted from industrialization. Instead, one may say that the emergence of industrial capitalism was politically possible only because it came with more inequality at the beginning. The relative diversification into commercial interests (i.e., the nonspecificity) of the old landowning class’s wealth also facilitated the transition to a liberal economic order. [For evidence on the diversity of economic investments of British landowners and how it helped to repeal the Corn Laws, see Schonhardt-Bailey (1991). For some confirmation of the Kuznets conjecture in Britain, see Lindert (2000). For partial confirmation in a cross-section of countries, see Barro (2008).]

Over time the new technologies of production became more widespread, either because they were actively promoted by capital owners interested in increasing the availability of complementary skills (Lindert 2004, pp. 87–122; Galor & Moav 2006) or because industrialists had no desire or capacity to block the access of low-income voters to those new technologies. In the context of the model, industrialists were either not the decisive voter or they were but their after-transfer income improved even after accepting the shock in place.

The spread of these new technologies of production had two related consequences: Incomes grew across the board and therefore strengthened the political capability of middle-income and low-income groups; and income compression increased (by the last third of the nineteenth century and at a faster pace since 1910), making democracy a more attractive political option than a narrow franchise (Boix 2003; see also Przeworski 2009 for an empirical confirmation of the same hypothesis, linking equalizing conditions and higher chances of democratization). The successive expansions of the franchise in Britain took place in lockstep with the evolution of the economy (Justman &
Figure 5

(a) Evolution of income shares in the United Kingdom and the United States. (b) Evolution of income shares in continental Europe.
Gradstein 1999). The expansion of the franchise then led to the expansion of public education, health spending, and direct transfers such as unemployment benefits and pensions, and to a further reduction in inequality (Boix 2001; Boix 2003, ch. 5; Lindert 2004).

The fragmented structure of the international system and the effect of interstate competition seem to explain well the progressive adoption of (or lack of opposition to) a modern regulatory structure. Reform happened first in North Atlantic economies with conditions similar to Britain’s (a “dormant” craft sector that could transform itself into a full-fledged industrial sector). It also took place in countries directly affected by an ever-growing security dilemma (Germany) or military defeat (Japan). However, interstate competition was not a sufficient condition of change. In highly unequal economies, that is, in economies in which elites’ wealth was highly specific and barriers to entry were central to the political-economic regime (the American South, Russia, China, most Latin American republics), reform thoroughly failed. The expected gains of reforms must not have matched the expected losses in the view of the elites. Once again, the closer any society was to the ancien régime “ideal type,” the more robust it turned out to be as a political equilibrium. Reactionary political economies only petered out as a result of domestic revolutions and, most often, of international war and invasions. Those places that were spared from interstate violence, such as Latin America, remained stuck in the old regime outcome.

As in Britain, all those followers who engaged in the construction of a liberal order and the process of economic development enjoyed a shift in their aggregate production function. Human capital became a central component of the economy. This resulted in a higher marginal product of labor, higher wages, and a generalized process of wage compression. This explains the secular decline in inequality across all Western countries in the past 100 years or so—without much need to resort to any kind of institutional story such as wage bargaining, electoral systems, and so on, except to account for (in relative terms) small cross-country differences. The share of income earned by the
top 10% in most advanced countries was on average about 20 percentage points higher in 1920 than in 1970. By contrast, the difference between maximum and minimum national values in 1970 was of about 5 percentage points. (Cross-national differences have increased in the past 20 years, however.) Moreover, with the exception of development models that strongly relied on state-led industrialization and the formation of tight-knit economic elites, such as Meiji Japan (Moore 1966), the process of economic development fostered the gradual liberalization of politics, which in turn led to the construction of a welfare state and growing equality of conditions.

**American Exceptionalism.** In almost all cases, economic development and political liberalization came together (probably with the latter—at least in the sense of broader political participation—following the former). By contrast, the northern half of the United States had relatively deep democratic institutions before industrialization started. **Figure 5a** shows how a relatively equal economy experienced growing inequality in wealth distribution (peaking at the turn of the century). Democratic institutions remained robust, although new franchise requirements were introduced and the electorate shrank a bit in the industrial North in the early twentieth century. In line with the distributive effects of regime type, human capital formation was stepped up, resulting in a strong leveling of wages across the labor market (see Goldin 1999 on long-run human capital formation in the United States). Democratic institutions, in interaction with the impact of massive immigration on the decisive voter’s income, were also likely to be behind the closing of American borders in the 1920s (see Mirilovic 2010 for a first cut on the impact of democracy on immigration policy). The ensuing fall in foreign-born population as a proportion of total population strongly correlated with a sharp decline in inequality (McCarty et al. 2006). A plausible interpretation of the historical data seems to confirm the idea that democracy’s institutions tend to generate policy responses to curb inequality-enhancing technological shocks.

**The Past 30 Years.** As is apparent in **Figure 5**, income dispersion has grown since the 1970s in some Western (mostly Anglo-Saxon) countries. Although the sources of the skill-biased shock are still under discussion, with the literature split between trade and strict technological factors, the diversity of response across OECD states seems to be related to their different institutional structures (mostly their wage-bargaining institutions) and partisanship.

This result is not at odds with the previous discussion. In the short run, certain institutional factors (such as wage bargaining and partisan governments) can reduce wage dispersion—at the price of high unemployment or subsidized employment through the public sector. But those institutions cannot explain long-run inequality because wage dispersion broadly tracks variance in marginal productivities in the labor market. Institutional conditions that create wage equality artificially lead to the misallocation of resources, and they end up leading to a suboptimal outcome in the long run. Of course they can sustain an artificially low level of wage dispersion (or, in fact, excessive income dispersion, as in authoritarian regimes) if they are complemented with measures that suppress the technological shocks that lead to differential incomes—for example, protectionism. But those measures are precisely what put the economy in the (old-regime) path of stagnation. Recent empirical studies confirm that wage-bargaining institutions explain levels of wage inequality after the 1970s but not before (Wallerstein 1999, Scheve & Stasavage 2009). In other words, Europe’s successful reduction of inequality over most of the twentieth century must be related to the construction of educated labor forces and a de facto absence of economic competitors abroad. But without those two conditions it will be unable to maintain a compressed wage structure without engaging in heavy spending and imposing new entry barriers.
CONCLUSION

Broadly speaking, the distribution of income in human societies has been characterized by the following patterns. Stateless, preagrarian societies displayed (and still display wherever they exist) relatively equal distributions of income and certainly a very low level of intergenerational transmission of wealth. The agricultural revolution and the concomitant formation of the state came hand in hand with the emergence of marked inequalities of income across individuals and over generations. The level of inequality has varied considerably across historical communities since the invention of agriculture. Income inequality has trended downward since the industrial revolution—although the degree of inequality has differed across countries and there have been some temporal reversals within the process of equalization.

After reviewing current economic and institutional models of the origins and evolution of inequality, this article first develops a theoretical model to explain the transition away from relatively equal preagrarian societies. A series of biased technological shocks generated inequality (within but mostly across human bands), made spontaneous cooperation unfeasible, and heightened violent conflict. The state of violent anarchy that followed opened the way to both political institutions and further shifts in income distribution. In response to the productivity growth, less productive individuals responded by creating protection institutions in exchange for systematic payments from more productive sectors. Alternatively, the latter could choose to devote part of their income to deter bandits. Both outcomes (“monarchical” and “republican”) had distinctive and sharp effects on the distribution of resources. Military technologies explained much of the variance in the prevailing political outcome.

Once political institutions were in place, income distribution depended on how technological shocks and/or policy regimes affected the net income of decision makers—directly, i.e., in the returns of their assets; and indirectly, i.e., after political institutions and tax rates, endogenous to the changing income distribution, were put into place. This framework may explain why inequality, authoritarianism, and economic stagnation go together in an ancien régime polity. It can be also employed to probe the causes of the technological breakthrough of industrialization and its political consequences. It is finally applied to discuss the limits of inequality under democracy and the relationship between democracy, technological shocks, and compensatory policies.

This more complex understanding of the sources of inequality turns out to affect our interpretation of the consequences of inequality for growth and development. Most of the recent work on growth and inequality claims that the latter has deleterious effects on economic development (Alesina & Rodrik 1994, Persson & Tabellini 1994, Perotti 1996). [By contrast, work done in the 1950s and 1960s had concluded that inequality covaried with growth, at least in its take-off phases (Kanbur 2000).] However, the empirical evidence that inequality is always a direct cause of underdevelopment is inconclusive. Several recent articles have shown that inequality has no effect on development or even raises growth rates (Li et al. 1998, Forbes 2000, Voitchovsky 2005, Barro 2008). A theoretically plausible interpretation of these results, given the political-economic framework presented here, is that the potential correlation between inequality and development is always conditional on the political and institutional causes behind a particular income distribution (Boix 2009). In ancien régime societies, wealth was based on the ownership of very specific assets—assets whose value relied on regulatory barriers sustained by the state, or assets (such as land) that could not be easily redeployed in new, more productive economic sectors. Because wealth was so specific to the sector in which it was deployed, the wealthy strata and the governing elite (which overlapped to a great extent) had very little interest in changing the status quo and accepting any technological breakthrough that could upset their dominant economic and political position. Within
that political and economic context, inequality and long-run economic stagnation (underpinned by a nondemocratic regime) ended up taking place simultaneously. That same pattern of differential access to the state and to regulation and wealth prevails in developing economies today and may well explain their high levels of inequality and meager growth. By contrast, in other instances, inequality covaries with growth. For example, the process of industrialization led, through technological change and factor shifting, to both higher incomes and more inequality, at least during its initial phase.

**DISCLOSURE STATEMENT**

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

**ACKNOWLEDGMENTS**

This article draws from work presented at the University of Virginia, the New York meeting of the “September Group,” the Santa Fe Institute, Princeton University, Stanford University, Yale University, the University of Chicago, and the London School of Economics. I am grateful for the comments of all their participants, in particular those of Alicia Adserà, Sam Bowles, Dan Gingerich, Stathis Kalyvas, Carol Mershon, and Thomas Romer. I thank Teppei Yamamoto for his research assistance.

**LITERATURE CITED**


