

COPYRIGHT NOTICE:

Peter Turchin & Sergey A. Nefedov: Secular Cycles

is published by Princeton University Press and copyrighted, © 2009, by Princeton University Press. All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher, except for reading and browsing via the World Wide Web. Users are not permitted to mount this file on any network servers.

Follow links for Class Use and other Permissions. For more information send email to: permissions@press.princeton.edu

Chapter 1

Introduction: The Theoretical Background

1.1 Development of Ideas about Demographic Cycles

The modern science of population dynamics begins with the publication in 1798 of *An Essay on the Principle of Population* by Thomas Robert Malthus. Malthus pointed out that when population increases beyond the means of subsistence, food prices increase, real wages decline, and per capita consumption, especially among the poorer strata, drops. Economic distress, often accompanied by famine, plague, and war, leads to lower reproduction and higher mortality rates, resulting in a slower population growth (or even decline) that, in turn, allows the subsistence means to “catch up.” The restraints on reproduction are loosened and population growth resumes, leading eventually to another subsistence crisis. Thus, the conflict between the population’s natural tendency to increase and the limitations imposed by the availability of food results in the tendency of population numbers to oscillate. Malthus’s theory was extended and further developed by David Ricardo in his theories of diminishing returns and rent (Ricardo 1817).

According to the Malthusian argument, the oscillation in population numbers should be accompanied by systematic changes in certain economic variables, most notably food prices. Fortunately, data on prices are reasonably abundant in historical sources, and it is possible to construct time series documenting price fluctuations over very long periods of time. Compilations of price trends appeared as early as the sixteenth century. For example, Ruggiero Romano (1967) reports that a time series of grain prices between 1500 and 1593 appeared in an appendix of *La Patria del Friuli Restaurata* by Jacopo Stainero, published in 1595 in Venice. The data on prices in medieval and early modern England were made available to historians by Thorold Rogers (1862). By the 1930s the empirical material had accumulated to the point where it became very clear that European prices had gone through a number of very slow swings between 1200 and 1900 (Simiand 1932, Griziotti-Kretschmann 1935, Abel 1980).

A most important and lasting contribution was Wilhelm Abel’s *Agrarkrisen und Agrarkonjunktur*, the first German edition of which was published in 1935. Abel compiled a rich data set containing time-series information about prices, wages, rents, and population movements in Western and Central Europe from the thirteenth to the twentieth centuries, ensuring

that the empirical importance of his work would remain high to this day. The most striking pattern to emerge was the wavelike movement of grain prices (expressed in terms of grams of silver). There were three waves or “secular trends” (Abel 1980:1):

1. An upward movement during the thirteenth century and early fourteenth century, followed by a decline in the late Middle Ages
2. Another upsurge in the sixteenth century, followed by a decline or apparent equilibrium (depending on the country) during the seventeenth century
3. A third increase during the eighteenth century, followed by irregular fluctuations during the nineteenth century, eventually converging to an early twentieth-century minimum

The twentieth century saw another (fourth during the last millennium) period of price inflation (Fischer 1996).

On the basis of the observed patterns, Abel argued that the fluctuations in the circulation of money could not adequately explain the long-term trends in the price of grain. By contrast, population moved more or less in the same direction as the food prices and in an inverse ratio to wages (Abel 1980:292–93). Abel concluded that the Malthusian-Ricardian theory provided a better explanation of the data than the monetarist theory. Furthermore, the Malthusian-Ricardian theory predicted that an increasing population would result in a specific progression of effects. Rents would rise first, with grain prices lagging behind rents, the price of industrial goods lagging behind grain prices, and workers’ wages bringing up the rear. The evidence showed that this was precisely what happened (until the whole system was dramatically changed in the nineteenth century).

Abel’s conclusions were soon supported and extended by other historians, with the most influential contributions made by Michael Postan, working in England, and Emmanuel Le Roy Ladurie, in France. In a talk given in 1950, Postan rejected a monetarist explanation of long-term price movements during the Middle Ages and firmly asserted the primacy of the demographic factor (Hilton 1985). Le Roy Ladurie was an even more consistent follower of Malthus. In *The Peasants of Languedoc*, first published in French in 1966, he argued that southern France went through a great agrarian cycle lasting from the end of the fifteenth century to the beginning of the eighteenth (Le Roy Ladurie 1974:289). Although Le Roy Ladurie did not completely ignore the social and political aspects of the cycle, his explanation of the causes underlying the cycle was firmly Malthusian. Speaking in 1973, he said, “it is in the economy, in social relations and, even more fundamentally, in biological facts, rather than in the class struggle, that we must seek the motive force of history” (quoted in Hilton 1985:4).

Such a radical Malthusian position could not but provoke a reaction from scholars working within the Marxist tradition. Although some Marxist historians doubted the very fact of a drastic and prolonged population decline from 1350 to 1450 (Kosminsky 1956), others accepted it but preferred to explain it as the “crisis of feudalism.” In an influential book first published in 1946, Maurice Dobb argued that the cause of the crisis was the inefficiency of feudalism as a system of production, coupled with the growing needs of the ruling class for revenue (Dobb 1963:42–47). The “feudal lust for expanded revenue” was a result of two processes: growth in the size of the parasitic class and the increasing extravagance of noble consumption. These two tendencies, working synergistically, resulted in an intensification of feudal pressure on the peasantry to the point where it destroyed the goose that laid the golden eggs. Dobb’s theory occasioned an extensive discussion. One interesting contribution to the theory was Paul Sweezy’s proposition that the growing extravagance of the feudal ruling class was a result of the rapid expansion of trade from the eleventh century onward, which brought an ever-increasing variety of goods within its reach (Sweezy et al. 1976:38–39). Thus, Sweezy sees the root cause of the fourteenth-century crisis in the impact of this exogenous force on the structure of feudalism (Sweezy et al. 1976:106).

Robert Brenner’s 1974 critique of Postan’s and Le Roy Ladurie’s theories might be regarded as a continuation of the Dobb-Sweezy debate of the 1950s (the “Brenner debate” papers are collected in Aston and Philpin 1985, Hilton 1985). Brenner did not deny that the Malthusian model had a certain compelling logic (Brenner 1985a:14). However, its attempt to explain long-term trends in economic growth and income distribution was doomed from the start because it ignored (“abstracted away”) the social structure, the most important part of which was the surplus-extraction relationship between the direct producers and the ruling class (Brenner 1985a:10–11).

One deficiency of the Malthusian theory, according to Brenner, was the empirical observation that different societies within Europe, starting from similar demographic and economic conditions obtaining after the Black Death, subsequently followed divergent trajectories. For example, serfdom completely disappeared from certain Western European countries (England, France) while making a strong comeback in Central Europe (Poland, Prussia). Thus, different property structures (the landholding system) and different balances of power (the cohesiveness and organization of the ruling class) could result in different paths followed by societies after the demographic catastrophe.

The second and even more damaging argument against the Malthusian model is the observation of continuous stagnation of most of the traditional European economies in the late medieval period (Brenner 1985a:18). For

example, the Black Death removed about one-third of the English population in the mid-fourteenth century, and by the end of the century the population had been further reduced to one-half of its 1300 peak. According to Malthusian logic, such a drastic population decrease should have led to higher agrarian productivity, low food prices and high real wages, and the resumption of vigorous population growth. Indeed, the dynamics of prices and wages were largely in line with the Malthusian predictions. Yet population stagnated for more than a century, with growth resuming only in the late fifteenth century. Brenner argued that such episodes of long-term stagnation could only be understood as the product of established structures of class relations (Brenner 1985a:18). A decline in the number of direct producers reduced the income of the lords. To maintain their income, the lords attempted to extract a greater amount from each peasant, as well as trying to dispossess one another (via brigandage and internal warfare). The result was the disruption of production, leading to a further demographic decline, rather than a return to equilibrium as the Malthusian model would predict (Brenner 1985b:224).

In their responses to Brenner's critique, Postan and Le Roy Ladurie were unable to effectively account for the prolonged post-Black Death depression phase within the Malthusian theory. Postan and Hatcher acknowledged the problem: "Indeed the reason why the recovery was so belated and so sluggish is still one of the incompletely resolved difficulties inherent in the medieval hypotheses Brenner disagrees with" (1985:69). On the other hand, the extreme version of the Marxist thesis (perhaps found in the purest form in Sweezy), which assigns class relations the all-determining role in the economic development of medieval and early modern Europe, would also fail to account for empirical facts. For example, such a purely class-struggle-based theory is unable to explain the secular cycles in population, prices, and wages, as well as why exploitation of peasants also fluctuated cyclically.

In the end, the critique of Brenner and certain others, most notably Guy Bois (1984), played a constructive role by pointing out that the Malthusian model neglects an important explanatory variable. What we need is a synthetic theory that encompasses both demographic mechanisms (with the associated economic consequences) and power relations (surplus-extraction mechanisms). In the dynamical systems framework, it does not make sense to speak of one or the other as "the primary factor." The two factors interact dynamically, each affecting and being affected by the other. We pursue this idea in the next section.

It is curious that both sides in the Brenner debate almost entirely ignored the role of the state. This omission is understandable. The Marxists tend to treat the state as merely a vehicle for conveying interests of the ruling class, while the Malthusians' focus has been on the economic variables.

There is, however, a significant movement among historical sociologists “to bring the state back in” (Skocpol 1979). States are not simply created and manipulated by dominant classes; they are agents in their own right, and they compete with the elites in appropriating resources from the economy.

Historians have long recognized that there were recurrent waves of state breakdown and political crises in European history: the “calamitous” fourteenth century (Tuchman 1978), the “iron century” of 1550–1660 (Kamen 1971), and the “age of revolutions” of 1789–1849 (Hobsbawm 1962). Each of these periods was preceded by a period of sustained and substantial population growth. In a pathbreaking book, Jack Goldstone (1991) argued that there is a causal connection between population growth and state breakdown. The seeds of this theory were already contained in the work of Malthus. Goldstone, however, does not argue that population growth is a direct cause of state collapse (in fact, he carefully distances himself from the strict Malthusian doctrine). Instead, population growth causes social crisis indirectly, by affecting social institutions, which in turn affect sociopolitical stability. For this reason, Goldstone refers to his theory as demographic-structural: *demographic* because the underlying driving force is population growth, *structural* because it is not the demographic trend itself that directly causes the state crisis but its impact on economic, political, and social structures (Goldstone 1991:xxvi). We discuss this theory in more detail in the next section, but here we should mention that the verdict on Goldstone’s work among historical sociologists has been highly positive (see, e.g., Collins 1993, Wickham-Crowley 1997, Li 2002).

To summarize, it is becoming increasingly clear to specialists from very diverse fields—demographers and historical economists, social historians, and political scientists—that European societies were subjected to recurrent long-term oscillations during the second millennium CE (Braudel 1988, Cameron 1989, Fischer 1996). Furthermore, the concept of oscillations in economic, social, and political dynamics was not discovered by the Europeans. Plato, Aristotle, and Han Fei-Tzu connected overpopulation to land scarcity, insufficient food supply, poverty, starvation, and peasant rebellions (Parsons 2005). The Chinese, for example, have traditionally interpreted their history as a series of dynastic cycles (Reischauer 1960, Meskill 1965, Usher 1989, Chu and Lee 1994). The fourteenth-century Arab sociologist Ibn Khaldun developed an original theory of political cycles explaining the history of the Maghreb (Inayatullah 1997). Are these phenomena, which at first glance seem very diverse, actually related? In this book we examine the hypothesis that secular cycles—demographic-social-political oscillations of very long period (centuries long)—are the rule rather than the exception in large agrarian states and empires.

1.2 A Synthetic Theory of Secular Cycles

The brief review in the previous section focused mainly on the controversies between advocates of various processes as dominant influences. In the heat of the debate, however, the opposing sides tend to simplify and caricature the views of each other. For example, it is clear that neither purely demographic nor purely class conflict explanations of secular cycles work very well when confronted with data. On the other hand, a synthetic theory that incorporates both of these (and some other) processes may provide us with a viable hypothesis that can be tested with data. The idea is that secular cycles can only be understood as a result of the interaction between several interlinked variables—economic (including demography), social structure (particularly, how the elites interact with the producing population and the state), and political (state stability or collapse). In the following paragraphs we sketch the outlines of such a synthetic explanation. Our explicit focus is on agrarian societies, that is, those in which more than 50 percent of the population (and typically above 80–90 percent) is involved in agriculture.

The Demographic Component

The demographic component of the theory is based very much on the original insights of Malthus and Ricardo, further developed by neo-Malthusians such as Le Roy Ladurie and Postan. The key variable is the population density in relation to the carrying capacity of the local region. The concept of *carrying capacity* was developed by ecologists in the context of the logistic model, invented by Paul Verhulst and popularized by Raymond Pearl (Pearl and Reed 1920). Carrying capacity is defined as the population density that the resources of the habitat can support in the long term (for an excellent discussion of human carrying capacity from an ecologist's point of view, see Cohen 1995). *Resources* usually refer to food, although in some environments the limiting resource may be the availability of water or fuel. Carrying capacity thus is an upper ceiling on population growth. From the point of view of economics, this limit arises because labor inputs into production suffer from diminishing marginal returns.

It is clear that the carrying capacity of a specific region is strongly affected by its physiographic features (the availability of land suitable for agriculture, water supply, soil characteristics, length of the growing season, and so on). It is also affected by year-to-year fluctuations in the temperature and the amount of rainfall, as well as by gradual changes in the climate. In other words, carrying capacity is a variable that changes in both space and time. Finally, and most important, carrying capacity is affected both by the existing level of agricultural technology and by how this technology

is employed. As Ester Boserup (1966, 1981) famously argued, population growth can have a positive effect on economic innovation.

Although Boserup is widely regarded as being anti-Malthusian, both her insights and those of Malthus can be comfortably combined within the same general theoretical framework (Lee 1986, Wood 1998). Thus, adverse effects of population growth on the standard of living can provide strong inducements for the adoption of new means of production. However, in agrarian societies, economic change can win only a temporary respite from marginal immiseration (Wood 1998, Clark 2007a). For example, a society that approaches the current limits of population growth can invest in clearing forests, draining swamps, irrigation, and flood control. All these measures will result in an increase in the carrying capacity. However, at some point there are no more forests to cut or swamps to drain, and if the population continues to grow, eventually it will again begin pressing against the Malthusian limits.

As population density approaches the carrying capacity, a number of related changes affect the society. There are shortages of land and food, and an oversupply of labor. As a result, food prices increase, real wages decline, and per capita consumption, especially among the poorer strata, drops. Economic distress leads to lower reproduction and higher mortality rates, resulting in a slower population growth. Should population density reach the carrying capacity, there would be just enough food to sustain and replace one individual; birth and death rates would equalize, and population density would be at an equilibrium. At least, this is what simple models such as the logistic predict; in actuality, other factors not taken into account by a purely demographic model would preclude the emergence of a stable equilibrium.

Population growth in excess of the productivity gains of the land has a fundamental effect on society's structures. The typical changes accompanying population growth are high rents and land prices, increasing fragmentation of peasant holdings or high numbers of landless peasants, and increased migration of landless peasants to cities. Urbanization (measured by the proportion of population inhabiting towns and cities) increases. Cheap labor results in a flowering of trades and crafts. The demand for manufactures increases, because the elites profit from high rents on land and lower labor costs. Increased urbanization and conspicuous consumption by the elites promote regional and international trade. The gap between the well-to-do and the poor grows. In rural areas overpopulation means that no food reserves are available in case of crop failure. Accordingly, years of poor harvest that would hardly be noticed in better times now result in significant mortality and, at worst, in catastrophic famines. Chronic undernourishment creates conditions conducive to the spread of epidemics.

The cities accumulate landless peasants and jobless artisans, who join the growing ranks of paupers and vagrants. Food riots and wage protests become frequent. Eventually, deepening economic misery leads to peasant and urban uprisings. However, as long as the elites are united and the state maintains control of the military, such popular uprisings have small chance of success. This fundamental point was recently reiterated by Jack Goldstone:

It is a profound and repeated finding that the mere facts of poverty and inequality or even increases in these conditions, do *not* lead to political or ethnic violence (Gurr 1980, Goldstone 1998, 2002b). In order for popular discontent or distress to create large-scale conflicts, there must be some elite leadership to mobilize popular groups and to create linkages between them. There must also be some vulnerability of the state in the form of internal divisions and economic or political reverses. Otherwise, popular discontent is unvoiced, and popular opposition is simply suppressed. (Goldstone 2002a)

Social Structure: Commoners, Elites, and Social Mobility

One important consequence of the law of diminishing returns is that the amount of surplus produced by cultivators is nonlinearly related to their numbers. Surplus is the difference between the total production and what is needed for subsistence (that is, the minimum amount of resources needed to support and reproduce each peasant household multiplied by the number of households). The amount of resources needed for subsistence increases linearly with population, while the total product grows slower than linearly as a result of the law of diminishing returns (figure 1.1a). As a result, at a certain critical population density, which we have defined as the carrying capacity, the two curves intersect. This is the point where the surplus becomes zero (and should population increase beyond the carrying capacity, the surplus becomes negative, with the consequence that peasant households do not get enough resources to reproduce themselves, and population density must decline).

The curve relating the amount of surplus produced to population density crosses zero both where population density equals zero and where it equals carrying capacity, and there is a hump somewhere between these two critical points (figure 1.1b). Thus, when population increases from a low level, initially the amount of surplus increases (more peasants means more surplus). At some intermediate density, however, the surplus reaches a maximum: this is where the effects of diminishing returns on labor inputs into agriculture begin to be felt. After that point, the surplus begins to decline.

The surplus produced by peasants is not made available to the elites (and the state) automatically; left alone, peasants would happily consume

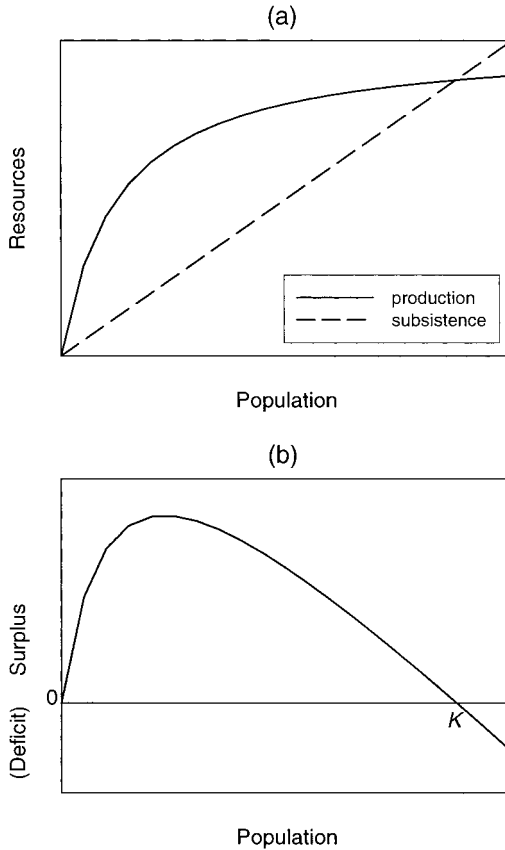


Figure 1.1 Effect of population growth on (a) total production, subsistence needs, and (b) the production of the surplus. K is the carrying capacity.

it themselves (or simply work less, “consuming” it as extra leisure time). How much of the production ends up in the hands of the elites depends on many economic and political factors. One important dynamic is that the elites are usually able to extract a larger amount of surplus during the late stages of population growth. Specific mechanisms depend on the land-holding system. For example, an oversupply of rural labor elevates rents and therefore increases a landowner’s profits. In a serfdom-based system lords can set the level of extraction almost arbitrarily high, because oppressed serfs have nowhere to flee—the whole surrounding landscape is at the saturation level, and the only alternative is the life of a vagabond or a bandit, which has always been brutish and short. Thus, most serfs have no realistic alternative to submission.

Prospects are equally bleak for free but landless laborers who must secure employment to support themselves and their families. Oversupply of labor leads to depressed wages and chronic unemployment or underemployment for a substantial part of population. On the other hand, employers, both rural and urban, profit greatly from this economic situation.

These considerations suggest that during the late stages of population growth, when commoners are already suffering from economic difficulties, the elites are enjoying a golden age. Both the reproduction of the existing elites and the recruitment of new elites from commoners will be fastest when the amount of extractable surplus is greatest. The expansion of elite numbers should take place during the “stagflation” phase (see below for the definitions of the phases of a secular cycle), when fast-rising prices and land rents offer the greatest opportunities for rapid accumulation of wealth by current and aspiring elites, and when state fiscal problems lead rulers to increase the sale of privilege and rank; both factors tend to accelerate social mobility into the elite ranks. As a result, the peak of elite numbers often lags behind that of the general population (the important exception of societies with widespread polygyny is discussed in section 1.3).

Such a happy state of events (for the elites) cannot continue for long. First, expansion of elite numbers means that the amount of resources per elite capita begins to decline. This process would occur even if the total amount of surplus stayed constant. But, second, as general population grows closer to the carrying capacity, surplus production gradually declines. The combination of these two trends results in an accelerating fall of average elite incomes.

The dynamic processes described above also have a sociopsychological aspect. During the good times the elites become accustomed to, and learn to expect, a high level of consumption (this is the “growing extravagance of noble households” of Dobb and Sweezy). An additional element, as pointed out by Sweezy, is the ever-increasing quantity and variety of goods available to the elites as a result of urbanization, the growth of crafts, and the expansion of trade (factors that are themselves a consequence of population growth). Modern studies of consumption level expectations suggest that people generally aim at matching (and if possible exceeding) the consumption levels of their parents (Easterlin 1980, 1996). Thus, what is important is not the absolute level of consumption but the level in relation to the previous generation. In other words, expected “living standard” is a culturally determined *inertial* variable (inertial because it changes slowly, on a generational time scale). If we can extrapolate results obtained by studying modern consumers to preindustrial elites (at least, this may be a reasonable working hypothesis), then we would predict that during the good times the elites would easily become accustomed to elevated levels of consumption, and this expansion would occasion little social comment. By contrast, should their level of consumption decrease in relation to the pre-

vious generation's, the elites would be expected to react vehemently to this development. This argument suggests there is no contradiction between the bitter critique of the elites for their luxurious and wasteful way of life by contemporary social commentators and the equally bitter complaining of the elites themselves about their poverty and indebtedness.

The deteriorating economic conditions of the elites during the late stagflation phase of the secular cycle do not affect all aristocrats equally. While the majority are losing ground, a few lineages, by contrast, are able to increase their wealth. The growing economic inequality results from the operation of what some sociologists call the "Matthew effect" (Merton 1968). Poor aristocratic lineages tend to get poorer because they attempt to maintain their elite status on an inadequate economic basis. This forces them into growing indebtedness, which eventually has to be addressed by selling some of their assets (such as land). A wealthier lineage, by contrast, can maintain the level of consumption necessary for preserving its elite status and have some resources left over to acquire land from its impoverished neighbors. As a result, the poor get poorer while the rich get richer. The same dynamic operates on peasants during the stagflation phase. During periods of economic hardship, poor peasants must sell land or starve. As a result, at the same time that the majority are sliding into absolute misery, a small percentage of thrifty, hardworking, or simply lucky peasants are able to concentrate increasing amounts of land in their hands. At some point, such successful peasants usually attempt to translate their wealth into higher social status. This demand for upward social mobility is an important factor contributing to the elite overproduction that develops towards the end of a prolonged period of demographic expansion.

During the stagflation phase, thus, economic inequality increases within each social stratum—peasants, minor and middle-rank nobility, and the magnates. Growing inequality creates pressure for social mobility, both upward and downward. Increased social mobility generates friction and destabilizes society. The growing gap between the poor and rich also creates a breeding ground for mass movements espousing radical ideologies of social justice and economic redistribution.

Dynamics of Surplus Extraction

The declining incomes of the majority of aristocrats have two important consequences: intensifying oppression of the peasants by the elites and increasing intraelite competition for scarce resources. The elites will attempt to increase the proportion of resource extracted from the producers by whatever means that are available to them, both economic and extraeconomic (coercive). Their success will depend on the structural characteristics of the society: the relative military strength of the elites with respect to the producers and the state, legal and cultural limits on surplus extrac-

tion, and so forth. If successful, elites may not only deprive the commoners of the surplus but may also cut into the subsistence resource, resulting in a negative growth rate for the commoner population. "Thus the lord's surplus extraction (rent) tended to confiscate not merely the peasant's income above subsistence (and potentially even beyond) but at the same time to threaten the funds necessary to refurbish the peasant's holding and to prevent the long-term decline of its productivity" (Brenner 1985a:31). It appears that this stage in the secular cycle may be what is known among dynamicists as a "bifurcation point," a point at which the system may follow one of several alternative trajectories. A classic example of such divergent trajectories is the disappearance of serfdom in post-medieval England and France and, during the same period, the rise of new serfdom in Prussia and Poland. Which of the alternative trajectories the system follows may depend on its structural characteristics or may be a result of a chance event. We are essentially rephrasing, in dynamical systems terms, the point made by Brenner in his critique of the Malthusian theory.

This thesis is illustrated by the recent study of Stuart Borsch (2005), which compared the effects of the Black Death in England and Egypt. In post-Black Death England wages rose, rents and grain prices dropped, unemployment decreased, and per capita incomes grew. Although the economic recovery of England occurred later than would be predicted by the Malthusian model, by the year 1500 it was in full swing. The consequences of depopulation in Egypt were profoundly different. Wages dropped, land rents and grain prices rose, and unemployment levels increased. No economic recovery was anywhere in sight by 1500. In fact, agricultural output declined between 1350 and 1500 by 68 percent. Borsch argues convincingly that the persistent stagnation of post-Black Death Egypt is explained by structural factors. After 1250 Egypt was ruled by a particularly cohesive and militarily capable group of elites: specialized slave-warriors known as Mamluks (as evidenced, for example, by their success at repelling the Mongol invasions in the second half of the thirteenth century). English peasants could resist elites by hiding in the hills and forests, of which there was an abundance in a depopulated England. Additionally, the longbow negated the advantage in military power usually enjoyed by the elites. By contrast, Egypt's narrow strip of arable land between uninhabitable desert left no room for evasive tactics. After the Black Death, Mamluks were able to use their tremendous coercive power to maintain the preplague level of resource extraction from a greatly diminished rural population. Extremely high levels of exploitation of individual peasants precluded any demographic revival. The system, thus, was caught in a "vicious equilibrium" that was apparently stable with respect to internal perturbations; it was finally destroyed by external conquest (the Ottomans in 1517).

The second consequence of plunging elite incomes is increased intraelite competition. The forms that this competition takes will depend (again) on the structural characteristics of the society. Probably the most important factor is the capability of the state to suppress overt violence. Here we consider the forms of intraelite competition in the presence of the state when internal order is maintained. The situation after the state collapses or is seriously weakened is considered later.

One recourse for elites facing declining incomes from agriculture was to seek employment with the state or church bureaucracy. Because training improved one's chances, a curious side effect of increased competition for such positions was the "credentialing crisis" (Collins 1979)—a rapid expansion of enrollments at the educational institutions (at least in those societies that offered formal training to aspirants for elite positions). Thus, we can use trends in higher education as an index of intraelite competition (Goldstone 1991:123). Another useful index of intraelite conflict is the level of civil litigation (Goldstone 1991:120).

Impoverished elites could also improve their incomes by attaching themselves to the retinues of powerful magnates. In fifteenth-century England this trend resulted in what is known as "bastard feudalism" (Dyer 1989:35). A large retinue was necessary to advance the lord's interests in government, litigation, and even civil war. However, limits on available land, civil and ecclesiastical offices, and royal patronage lead to increasingly polarized factional battles between patron-client groups for available spoils (Goldstone 1991:119). As a result, the elites tend to lose their unity and split along numerous fission lines: new elites versus old, one religious faction against the other, regional elites against the center, and so on. Because there are not enough resources for everybody, certain segments of elites, or groups aspiring to elite status, inevitably end up as the losers. We refer to them as the counterelites, or dissident elites. Usually, the counterelites do not constitute a true sociological group, because there is little that unifies them apart from hatred for the existing regime and a burning desire to bring it down. Incidentally, we are not implying here that the motivations of the counterelites are purely economical. The late stagflation phase, as we argued above, is typically characterized by a harsh oppression of the productive segments of the society and extreme social inequality, offering ample ideological justification for revolutionary action.

State Breakdown

Social trends resulting from demographic growth—declining surplus production, popular immiseration, and intraelite competition—have a profound impact on the ability of the state to maintain internal order, or even to survive (Goldstone 1991). Population growth leads to expansion of ar-

mies and bureaucracies, resulting in rising state expenditures. An increased number of aspirants for elite positions puts further fiscal strain on the state. Thus, states have no choice but to seek to expand taxation, despite resistance from the elites and the general populace. Yet the amount of surplus production declines (as discussed in the previous section), and the state must compete for this shrinking surplus with increasingly desperate elites. As a result, attempts to increase revenues cannot offset the spiraling state expenses, and even though the state is rapidly raising taxes, it is still headed for fiscal crisis. Note that declining *real* revenues may be masked by persistent price inflation, and it is therefore important to express all fiscal flows in real terms.

As we discussed in the previous section, population growth leads to rural misery, urban migration, falling real wages, and an increased frequency of food riots and wage protests. After a certain lag time, the negative effects of population expansion begin to affect the elites, who become riven by increasing rivalry and factionalism. Another consequence of rapid population growth is the expansion of youth cohorts. This segment of the population is particularly impacted by lack of employment opportunities. Finally, growing economic inequality, elite competition, and popular discontent fuel ideological conflicts. For example, in early modern Europe, dissident elites and dissatisfied artisans were widely recruited into heterodox religious movements.

As all these trends intensify, the end result is state bankruptcy and consequent loss of the military control, elite movements of regional and national rebellion, and a combination of elite-mobilized and popular uprisings following the breakdown of central authority (Goldstone 1991:25). Internal war among political factions is only one aspect of increased interpersonal violence. A breakdown of social order is also accompanied by increased banditry, homicides, and other kinds of violent crimes. On the ideological level, the feeling of social pessimism is pervasive and the legitimacy of the state authority is at its lowest point. The society approaches a condition that may appropriately be called “Hobbesian” (Hobbes himself lived during such a period). We refer to these conditions collectively as high sociopolitical instability.

The Effect of Sociopolitical Instability on Population Dynamics

In the previous sections we focused on the manifold effects of population growth on various structures of the society, including a bundle of variables that we call sociopolitical instability. Here we consider the feedback effect: how does instability affect population dynamics? We can envision two general (and, actually, interrelated) ways: by affecting demographic rates and by affecting the productive ability of the society (Turchin 2003b:120–21).

Most obviously, when the state is weak or absent, the populace will suffer from elevated mortality due to increased crime, banditry, and internal warfare (civil war). External war may also play a role. Although external warfare between states has been a constant feature of agrarian states, its effect on demography should change with the phase of the secular cycle. When the state is strong, warfare is directed outward, and the areas that suffer most are the state frontiers, as well as areas outside, which are targeted for conquest. Collapse of the state and the ensuing civil wars reduce the resistance of the society to external invasion. As a result, internal warfare and external invasions by groups ranging from small bands of raiders to rival great powers can become hard to separate (this is, for example, what happened during the Hundred Years' War in France). Warfare has also an indirect effect on mortality, because movements of rebel or invading armies spread epidemics.

The times of trouble also cause increased migration: refugees flee from war-afflicted areas or areas whose productive potential has been destroyed. Migration has several effects. First, it can lead to emigration (and we can simply add that to mortality). Second, people on the move cannot afford to have children. Thus, birth rates decline. Third, migration leads to epidemics. Increased vagrancy spreads the disease by connecting areas that would stay isolated during better times. For example, in Ireland during 1810–44 (the period just before the Great Famine), harvests failed or partially failed in fifteen years out of thirty-five. These failures did not lead to starvation, but they were followed by outbreaks of “famine fevers”—typhus, dysentery, scurvy, cholera—which were spread throughout the isle by beggars and vagrants seeking charity and employment (Grigg 1980:138).

Additional factors facilitating the spread of disease are the movements of armies and the expansion of international trade. The latter factor should be qualified by noting that international trade expands in the precrisis period (stagflation phase) and then gradually declines after the society has descended into anarchy. Thus, the rise of widespread epidemics—pandemics—is most probable during the late stagflation phase. In fact, the arrival of a pandemic is one of the most frequent triggers of the demographic-structural collapse.

On a more local scale, vagabonds and beggars aggregate in towns and cities, increasing their population size. This may tip the local population density over the epidemiological threshold (a critical density above which a disease spreads and below which it dies out).

Finally, political instability causes lower reproduction rates, because personal consumption plummets as a result of lowered production capacity. In the absence of organized ways to store surplus, peasants are unable to weather short-term subsistence crises. What stores are accumulated by individual households are easy prey to the marauding armies and other pred-

ators. In addition, during times of uncertainty people choose to marry later and to have fewer children. Incidentally, people's choices about their family sizes may be reflected not only in birth rates but also in the rates of infanticide. Thus, family limitation practices may be disguised as increased infant mortality.

The second and perhaps even more important effect of sociopolitical instability is on the productive capacity of the society (the carrying capacity). Vigorous states often invest in increasing the agricultural productivity by constructing irrigation canals and roads, implementing flood control measures, clearing land from forests, organizing the colonization of underpopulated regions, and so on. The end result of these measures is mainly an increase in cultivated area, although some measures also increase the productivity of land.

The other general mechanism is that the state offers protection. In a stateless society, people can live only in natural strongholds or in places that can be made defensible, such as walled cities. For example, at the height of the Roman Empire the overwhelming majority of the Italian population was to be found in the lowlands, where the most productive land was concentrated. After the collapse of Rome, settlements were moved to hilltops (Wickham 1981). An even more striking illustration comes from the Wanka hill fort chiefdoms in the Mantaro Valley of Peru (Earle 1997). Prior to Inka pacification of the region, the Wanka lived in crowded hilltop fortresses. After the conquest, the population moved down to lower elevations, where the best agricultural land was located. As a result, the diet and life span of both elite and commoner were dramatically improved (Johnson and Earle 2000:327).

The third example comes from *Histoire de Charles VII* by the Norman bishop Thomas Basin, who described northwestern France during the 1420s, after a particularly virulent outbreak of the Hundred Years' War:

a state of devastation such that from the Loire to the Seine, and from there to the Somme, the peasants having been killed or run off, almost all fields were left for a number of years not only uncultivated, but without people. . . . All that could be cultivated at that time in that region was only around and inside towns or castles, close enough so that, from the top of the tower or watchtower the eye of the lookout could perceive the attacking brigands. Then, with the sound of a bell, or horn, or some other instrument, he gave all those working in the fields or vineyards the signal to withdraw to the fortified place. (quoted in Dupâquier et al. 1988a:368)

In other words, lack of effective suppression of internal violence by the state imposes a "landscape of fear," in which a large proportion of agriculturally suitable lands is abandoned because they are too far from a place of

security. By contrast, the strong state protects the productive population from external and internal (banditry, civil war) threats, and thus allows the whole cultivable area to be put into production.

Elite Dynamics during the Depression Phase

Sociopolitical instability affects elite numbers in a fashion that is similar to its effect on commoners, although the relative importance of specific mechanisms can be quite different. Thus, the elites may be little affected by subsistence crises. They also tend to escape more lightly the effect of epidemics. This is partly due to their better nutrition and the likelihood of getting better care during disease, but even more important is their higher mobility. Urban elites could withdraw to their country estates at the first sign of incipient epidemic (as in Boccaccio's *Decameron*). Higher nobility with estates in multiple provinces could similarly avoid an epidemic striking a particular region.

On the other hand, by virtue of their more active participation in politics, the elites ran a much higher risk of violent death. The death toll in some conflicts was extraordinary. For example, Dupâquier et al. (1988a:342) quote an estimate by Philippe Contamine that around 40 percent of the French elite may have been slaughtered in the Battle of Poitiers (1356), and the same proportion at Agincourt (1415). During the Wars of Religion in the late sixteenth century, 20,000 Huguenots were killed in just one day, the St. Bartholomew's Day massacre (Kamen 1971:39).

Loss of life or elite status could also result from state purges. For example, the first Ming emperor purged 100,000 Chinese officials (Tignor et al. 2002:62). Sulla's proscriptions eliminated a third of the Roman ruling class, senators, and another third was eliminated by proscriptions following Caesar's death (see chapter 6).

A much less spectacular but perhaps ultimately more important process reducing the elite numbers is downward mobility. The plunge in elite incomes, which begins in the precrisis period and is greatly exacerbated by the general population decline, affects most strongly the status of the lowest noble stratum. A specific example is given by Christopher Dyer for late medieval England. An esquire or gentlemen living on £10–20 a year who was employing only three servants and lived in one house, and whose meals were devoid of much luxury in terms of wine and spices, had little room to maneuver when his income plunged by up to 50 percent in the mid-fifteenth century: "They must have cut back, or even cut out completely, their occasional wine-bibbings, and avoided travel whenever possible, but too many economies of this kind might force them to drop out of the aristocracy and accept yeoman status" (Dyer 1989:108).

In summary, a number of social mechanisms exist by which elite surpluses can be reduced: (1) deaths resulting from civil war, (2) deliberate purges of elites by new rulers, (3) limitations imposed on heir production (celibacy, primogeniture), (4) downward social mobility, voluntary or forced by the state, (5) increased material resources resulting from conquest or improvements in agricultural productivity, and (6) the development of a new political order that directs a greater share of resources to the elites. Several such mechanisms are usually operating in combination; the specific mix depends on cultural peculiarities of societies and historical accidents.

End of Instability and the Beginning of the New Cycle

Because the three main factors driving the rise of sociopolitical instability are general overpopulation, elite overproduction, and state insolvency, all these trends must be reversed before the disintegrative phase can end. Such trend reversal can occur in a variety of ways, depending on the characteristics of the society, its geopolitical environment, and various other exogenous factors. As a result, the last stages of the secular cycle are particularly rife with bifurcation points, and the sociopolitical trajectory can behave in a very nondeterministic fashion.

The problem of overpopulation is usually “dealt with” during the crisis phase. One of the most common proximate mechanisms of population collapse is disease, but not all population declines are accomplished by catastrophic epidemics. Prolonged periods of civil war can also cause drastic drops in population levels, although typically requiring more time. Finally, the external conquest of a disunited society often results in a demographic catastrophe.

An alternative to population collapse is an increase in the carrying capacity—after all, overpopulation results not from the absolute numbers being too large but from too high a population density in relation to the carrying capacity. The carrying capacity can increase as a result of technological progress. This is probably what happened in early modern England. During the crisis of the seventeenth century, the English population hardly declined, while the average yield of grain per acre probably doubled. The end result was a twofold decline in the population pressure on resources.

The carrying capacity may also increase as a result of the conquest of new underpopulated territories. An example is the conquest of the Kazan and Astrakhan khanates by Muscovy in the sixteenth century, which opened up vast areas along the Volga for Russian colonization during the succeeding centuries. Theoretically, the carrying capacity can also increase as a result of a substantial amelioration of the climate, although at this point

we cannot point to a well-documented, convincing example of this mechanism in action.

The processes leading to the reduction in the elite numbers and appetites were discussed in the previous section. The manner in which elite overproduction is abated depends very much on the military strength of the aristocracy. A nonmilitarized ruling class can be expropriated en masse by warlords, such as rebel generals or even peasant bandits. A rapid and comprehensive elite turnover results in a relatively short period of sociopolitical instability that follows state collapse. This is apparently what happened on several occasions during the Chinese imperial period, where the ruling class was dominated by the literate administrative elites rather than by military specialists. A rapid elite turnover can also result when there is a ready external source of potential elites, as was the case in the Maghreb described by Ibn Khaldun (we discuss Ibn Khaldun cycles in the next section).

A ruling class that enjoys a preponderance of military power over both internal and external rivals can be reduced only by internecine fighting between various elite factions. This can result in very prolonged periods of sociopolitical instability, or “depression” phases, in our terminology.

Thus, for a new secular cycle to get going, the pressures of the general population on resources and of the elites on commoners must be substantially reduced from their precrisis levels. There is also a third condition. Not all societies are capable of the broad-scale cooperation that is required to construct a functioning state, and some societies with a previous imperial history can also lose this ability with time (Turchin 2003b, 2006). Thus, it is entirely possible for the civil warfare to gradually die out but a centralizing, integrative trend nevertheless failing to take hold. In this case, the area in question may persist indefinitely (or until it is conquered from the outside) in a fragmented state as a collection of small-scale polities. The potential explanations of this failure to build a functioning state lie beyond the scope of our book. Here we simply indicate that it is yet another possible bifurcation point.

Phases of the Secular Cycle

Oscillatory dynamics do not go through truly discrete phases with clearly marked breakpoints, but for convenience in talking about each secular cycle, we need to divide it in phases. Our classificatory scheme is given here with the understanding that transitions between phases are rarely abrupt, so that any particular year that we designate as an end to one phase and the beginning of another is to some degree arbitrary (for this reason, we usually round the date to the nearest decade).

Most broadly the cycle can be divided into two opposite trends. In the literature these are sometimes called the positive “A phase” and the nega-

tive “B phase,” but we prefer the more descriptive terms *integrative* and *disintegrative* trends. Politically the integrative phase is characterized by a centralizing tendency, unified elites, and a strong state that maintains order and internal stability. Internal cohesion often results in the vigorous prosecution of external wars of conquest, which may result in the extension of the state territory (assuming there are weaker neighbors at whose expense the state can expand). The disintegrative phase, by contrast, is characterized by a decentralizing tendency, divided elites, a weak state, and internal instability and political disorder that periodically flare up in civil war. External wars of conquest are much more difficult to prosecute during the disintegrative phase. If they happen, they usually take place during the intervals between civil wars and at the expense of equally weak opponents. More frequently it is the external enemies that profit from the internal weakness of the state and society, resulting in an increased frequency of raids, invasions, and loss of territory.

The population tends to increase during the integrative phase and decline or stagnate during the disintegrative one. Climatic fluctuations, epidemics, or being overrun by an external enemy can cause short-term (if significant) population losses. However, vigorous population growth resumes as soon as such exogenous forces stop acting. During the disintegrative phase, by contrast, population losses due to epidemics, famines, or wars are not made up by sustained population growth. Even when the proximate Malthusian forces (epidemics, famines, and wars) are in abeyance, the population often fails to increase, despite being much below the carrying capacity.

It is useful to further divide the broad integrative and disintegrative periods into subphases. Population growth is particularly vigorous during the first, *expansion* phase of the integrative trend. This is a time of relatively stable prices and modest real wage declines (if any). However, as the population density begins to approach the limits set by the carrying capacity, price increases or wage declines accelerate—this is the “stagnation” or “compression” or even more descriptively *stagflation* (stagnation plus inflation) phase. Although the majority of commoners experience increasing economic difficulties during the stagflation phase, the elites enjoy a golden age, and their numbers and appetites continue to expand.

The stagflation phase (and the overall integrative trend) is succeeded by a general crisis. Whereas expansion grades smoothly into stagnation, the transition between stagflation and crisis is often (but not always) abrupt. Discrete events signaling the arrival of crisis can be pandemics, extreme episodes of famine, or state collapse followed by intense civil war (or any such events in various combinations). The crisis phase in our terminology is not a discrete, brief event (which is one meaning of the word *crisis*) but an extended period that can last for one or more human generations. The

decline of population numbers during a crisis results in a situation of plentiful per capita resources. However, this does not necessarily end the disintegrative trend, because there are usually too many elites and elite aspirants, and intraelite conflict continues to generate internal instability. Thus, the crisis grades smoothly into a *depression* phase, characterized by endemic civil warfare. The population may grow during the intervals between intense civil wars, but such increases typically do not last and are followed by declines (although not as a catastrophic as those typical of the crisis phase). The depression phase ends when the ranks of elites are pruned by internal conflict to the point where the disintegrative trend can reverse itself, and a new secular cycle begins. Alternatively, if no functioning state can get going, then the depression phase grades smoothly into an *intercycle* of indeterminate length.

We wish to emphasize again that the classificatory scheme we propose is an ideal type. It is helpful to be able to indicate the rough state of the dynamical system with a single word. However, there is considerable variation in the trajectories followed by actual societies. Thus, the boundary between various phases should be taken as “fuzzy” rather than “hard.” An early reader of the book manuscript even suggested that, instead of dividing the timelines of the societies that we study too neatly, we could allow phases to overlap. There is some merit to this suggestion, because different phases are dominated by different kinds of social processes, and these processes often overlap in time. For example, the onset of political crisis does not always coincide with the shift from population growth to population decline, and therefore the dating of the stagflation-crisis transition may be problematic. In the end, we chose to stay with nonoverlapping phases, because doing otherwise would be too confusing to our readers. But we do not impose these discrete phases on each case study in a procrustean manner.

1.3 Variations and Extensions

Factors Affecting Characteristic Lengths of Secular Cycles

Our exposition and illustration of the general theory of secular cycles in section 1.2 was Western European-centric, but the theory should, in principle, be applicable to any agrarian society. In this section we discuss how certain structural and cultural characteristics of societies should affect the demographic-structural dynamics, with a focus on one of the most important characteristics of oscillatory dynamics, the average period of a cycle.

Secular cycles are not periodic in the strictly mathematical sense, in which each succeeding cycle repeats exactly the preceding one. Although the secular rises and falls are generated endogenously by interactions be-

tween various components (subsystems) of the agrarian state, macrosocial dynamics of agrarian states cannot be strictly periodic. There are at least three reasons for that. First, nonlinear dynamic feedbacks can in theory generate not only strictly periodic (cyclic in the mathematical sense) dynamics but also aperiodic chaos—erratic-looking behavior that is nevertheless produced entirely by internal, endogenous reasons. The more complex the system (the more components it has) and the more nonlinear the interactions between the components (such as the presence of threshold responses), the greater is the likelihood that its dynamics will be characterized by sensitive dependence, the hallmark of chaos. Social systems are complex and feedback loops are nonlinear, so the possibility of chaos cannot be discounted (Turchin 2003b).

Second, the dynamics of agrarian states are affected not only by their internal workings but also by exogenous forces, such as changes in their geopolitical and ecological environment. Exogenous factors, unlike endogenous ones, are those that are not part of feedback loops (Turchin 2003a): they affect societal dynamics but are not themselves influenced by societal dynamics.

Finally, individuals possess free will and can act in unpredictable ways. In principle, even the act of a single person, if it takes place in the right place at the right time, may be able to influence the trajectory of a whole society. For lack of better theoretical approaches, we can model actions of individuals at the microlevel as a stochastic process, a kind of Brownian motion that also results in erratic, unpredictable changes in the macrosocial trajectory.

For all these reasons, we do not expect a strict periodicity in secular dynamics. Instead, dynamics should have an average period, a characteristic time scale, with a substantial degree of variation around this average. The mean period of a single—boom and bust—secular cycle is determined by the characteristic lengths of its phases, which in turn depend on various social, economic, and political parameters. Thus, the typical length of the expansion phase is primarily determined by (1) the per capita rate of population increase and (2) the population density in relation to carrying capacity at the beginning of the cycle. For example, if population grows at the rate of 1 percent per year, it takes seventy years for it to double. This is not a bad estimate of a typical expansion phase.

Expansion phases are also affected by geopolitical environment. States enjoy the greatest ability to mobilize the society for a war of external conquest during the middle parts of integrative secular trends. Abnormally long expansion phases result from successful territorial conquest, especially when it is accompanied by colonization of conquered territories, which serves to reduce population pressure in the metropole.

The length of the crisis phase is much less predictable, because while there is a definite biological limit on how fast a human population can grow, there is no comparable limit on how fast it can decline. Depending on the agent of change, population can decline very rapidly, as in a pandemic, or more slowly, due to incessant civil warfare. Furthermore, pathogens afflicting historical populations varied in their lethality. A relatively mild pathogen could drive population down slowly (perhaps as a result of recurrent epidemics), resulting in a long decline phase. A severe epidemic, on the other hand, would lead to a very short period of drastic population decline, and also to a deeper degree of social disintegration and longer depression phase (as happened in post-Black Death Europe).

The characteristic lengths of the stagflation and depression phases depend more on the state and, particularly, on elite dynamics than on what the general population does. In particular, the military strength of the elites has a large effect on the length of the depression phase, or even if there is such a phase at all. Models tailored to the characteristics of Western European societies (largely monogamous elites enjoying a preponderance of military power over their internal and external enemies) suggest that the typical periods of secular cycles in these societies should lie in the range of two to three centuries (Turchin 2003b:138).

Ibn Khaldun Cycles

A very different situation obtains in certain Islamic societies. The paradigmatic example is the sociopolitical dynamics in the medieval Maghreb, brilliantly described by Ibn Khaldun (1958). From the point of view of the demographic-structural theory, the Maghrebin states differ from Western European states in two important respects: (1) these Islamic societies permitted polygyny and (2) there was a ready source of militarily powerful counterelites nearby.

Polygyny is important because the number of wives is the most significant predictor of male reproductive success in humans (Betzig 1986). Because aristocratic males could afford to support several wives and concubines, the rate of elite population growth in Islamic societies was (and is today) much greater than that for elites in Christian societies. It is true that some degree of elite polygyny was practiced in Western Europe, where aristocrats often increased their biological fitness by having multiple mistresses and then acknowledging their bastards. Nevertheless, the fact remains that the biological reproduction rate of Islamic elites was several times higher than that of Christian elites.

The second factor is the location of Maghrebin societies in the rather thin strip of arable land squeezed between the Mediterranean Sea and the desert. The “desert” (or rather dry steppe and semi-desert zone between

the agrarian societies of the Mediterranean littoral and the extremely arid central regions of the Sahara) was occupied by nomadic pastoralists, primarily the Berbers. These desert chiefdoms were not a significant military threat as long as the agrarian states maintained their internal cohesiveness. But as soon as a Maghrebin society experienced state collapse, it became extremely vulnerable to conquest from the desert.

When demographic-structural models are modified to account for these two factors they exhibit very different dynamics (Turchin 2003b). High reproductive rate of the elites means that they increase much faster than the general population. Elite numbers, in fact, increase so rapidly that the commoner overpopulation plays a much lesser role or even no role in bringing about the state collapse. As a result, the integrative trend of the secular cycle is over much faster than in the standard model, developed for the Western European situation. Once the collapse occurs, there is usually no lengthy depression phase, because it does not take much time to organize a coalition of desert tribes to pick up the pieces and establish a new dynasty.

As a result of a shortened integrative trend and a missing depression phase, models predict a much faster secular cycle for Maghrebin-type societies, on the order of one century, rather than the two to three centuries for Western European states. This prediction is in agreement with the observation of Ibn Khaldun that the dynastic cycle in the Maghreb extends, on average, over four generations (a generation time in humans is typically twenty to thirty years). Note that this is a true theoretical prediction: models were not fitted in any way to the Maghrebin data. The shorter cycle period follows directly from the structural assumptions of the models of a faster elite reproductive rate and rapid elite turnover after state collapse.

Not all Islamic polities are predicted to exhibit Ibn Khaldun cycles. The key parameter, identified by the theory, is the rate of growth of elite numbers. Islamic societies that controlled the elite growth rates in one way or another are predicted to exhibit slower cycles, with periods similar to those observed in Western Europe. For example, in the Ottoman Empire the sultans had access to an essentially unlimited supply of wives and concubines. However, when the old ruler died, only one son was allowed to replace him; all others were killed. Furthermore, top levels of bureaucracy and army leadership were recruited not from native elites but by means of *devshirme*. In other words, the state, not biology, controlled the size of the high-ranking elite stratum. Only lower-rank landed elites were permitted to increase “biologically,” and, being not very wealthy, they could not afford too many wives. As a result, we can predict that secular cycles in the Ottoman polity should be much longer than those in the Maghreb.

An even more extreme case is the Mamluk polity in medieval Egypt. Its ruling class was recruited entirely from the slave markets. Children of Mamluks could not be Mamluks, and thus automatically dropped out of the ruling class. In principle, this arrangement should have stopped dead the Ibn Khaldun's dynamic and, barring exogenous perturbations, should have led to a stable equilibrium.

The Fractal Nature of Historical Dynamics

In general, different social processes operate at a variety of temporal scales. The shorter scales include daily, weekly, monthly, and annual cycles. Beyond that we have human generations, processes occurring on the time scale of centuries (including secular cycles), and longer-term phenomena such as social and biological evolution. As an example we can consider the stock market, as measured by the Dow-Jones Industrial Average (DJIA). The DJIA fluctuates on a variety of scales: daily (because the stock exchange shuts down at night), weekly (no activity on weekends), annual (fiscal year accounting affects trader behavior), multi-annual (business cycles), and multidecadal (the Kondratieff cycle, although not all economists accept the reality of such long cycles). The DJIA trajectory looks "fractal" because the amount of fluctuation depends on the time scale at which the trajectory is viewed.

If we are interested in understanding the effect of the business cycle on stock prices, we really do not care about short-term fluctuations. We certainly should ignore price movements within a single day, and probably even within a week. Thus, the time series with which we would want to investigate multi-annual oscillations would probably use DJIA values averaged for each week. Averaging is the simplest kind of smoothing, so what we have done is essentially smoothed away all "uninteresting" short-term fluctuations—uninteresting, that is, from the point of view of the main question of analysis. On the other hand, if we want to know how holiday periods affect stock price movements, we would certainly want to retain within-week fluctuations, and perhaps go down to hourly movements (to see how trading patterns behave during the short preholiday days). Now the variation due to the business cycle becomes a nuisance, and it might be a good idea to remove the effect of multi-annual and longer-term fluctuations by detrending. The point is that different questions require approaching an analysis at different time scales.

Turning now to population dynamics, we observe that population changes also occur on a variety of scales: monthly (female menstrual periods), yearly (subsistence and epidemic cycles), generational (somewhere between two and three decades), and secular (one, two, or three centuries, according to the theory of secular cycles). If we are interested in the dynam-

ics of childhood diseases, then the appropriate time scale would be weeks or months, to capture the within-year course of each epidemic (the incidence of measles, for example, begins to increase after children are brought together at the beginning of the school year, gradually building up to a peak in winter).

If we want to understand how secular cycles unfold, on the other hand, we certainly do not need to know how mortality fluctuates on a weekly or monthly time scale, or that there may be a deficit of births nine months after Lent as a result of devout Christians avoiding sexual intercourse. All such within-year or even year-to-year fluctuations are irrelevant for the purposes of our investigation. The appropriate time step is one human generation, and we need to average over smaller-scale fluctuations. We also need to do something about very long trends driven by social evolution. This requires some kind of removal of millennial trends (Turchin 2005:153), for example as was done for the early-modern English population (see appendix to chapter 3). By smoothing within-decade fluctuations and removing millennial trends, we retain two temporal scales of interest. The longer one is the average period of the secular cycle—this is what needs to be explained. The shorter one is the human generation time—this is the time step of the dynamical process that is postulated to be the explanatory mechanism of secular cycles.

It is important to remember that population numbers are a dynamic variable that has a lot of inertia on temporal scales shorter than a human generation. This is particularly true with respect to population increase: it can occur only slowly as babies are born and raised to enter the adult population. Even under ideal conditions, the human population needs at least one generation to double. On the other side of the demographic balance, mortality, it is theoretically possible for a population to collapse to a very low level (or even to go extinct) in a very short time. However, most typically annual variation in death rates, due for example to crop failures, can be quite substantial but is largely smoothed out—buffered—at the level of total population numbers.

This buffering ability of total population numbers is important in understanding how climate variability affects population dynamics. Annual variation, even if quite extreme, may have little effect on population change. If the population is well below carrying capacity, peasants may have sufficient stores to weather a year or two of bad crops without any demographic effect. In contrast, a long-term cooling, even if by less than one degree centigrade, may have a much more substantial effect on population dynamics by lowering carrying capacity. (This argument is just an illustration of why temporal scales are important; in the real world, the effect of climate change depends on the phase of the cycle, the alternative crops that peasants can switch to, and many other factors.)

Generation Cycles

The preceding discussion should make it clear that we are far from adopting a monocausal view of human history. The main hypothesis of this book is that demographic-structural processes are very important in historical dynamics, but we would be the last to argue that they are the only thing that goes on. However, it is not a good research strategy to include everything one can think of in the model. The history of science shows, over and over again, that an attempt to incorporate too many explanatory factors into theories is self-defeating. As Albert Einstein once said, a theory should be as simple as possible, but no simpler than that.

One particular process that is not part of the demographic-structural theory but has to be taken into account when studying secular cycles is the “fathers-and-sons” dynamic (Turchin 2003b, 2006). This mechanism operates during the prolonged disintegrative secular trends that are characteristic of secular cycles in Europe. The empirical observation is that disintegrative trends are not periods of continuous civil war; in fact, there are periods when sociopolitical instability is particularly high, interspersed with periods of relative pacification.

To illustrate this dynamic, during the disintegrative trend of late medieval France (“the Hundred Years of Hostility”), good reigns alternated with bad ones. The reign of John II (1350–64) was a period of social dissolution and state collapse, while that of his son Charles V (1364–80) was a time of national consolidation and territorial reconquest. The next reign, that of Charles VI (1380–1422), was another period of social disintegration and collapse. It was followed by a period of internal consolidation and national resurgence under Charles VII (1422–61), which finally lifted France out of the late medieval depression. This is a general dynamical pattern of alternation between very turbulent and relatively peaceful spells that is observed again and again during the secular disintegrative phases. A possible explanation of such swings in the collective mood lies in the social psychology.

Episodes of internal warfare often develop in ways similar to epidemics or forest fires. At the beginning of the conflict, each act of violence triggers chains of revenge and counter-revenge. With time, participants lose all restraint, atrocities become common, and conflict escalates in an accelerating, explosive fashion. After the initial explosion, however, violence drags on and on, sometimes for decades. Sooner or later most people begin to yearn for the return of stability and an end to fighting. The most psychopathic and violent leaders are killed off or lose their supporters. Violence, like an epidemic or a forest fire, “burns out.” Even though the fundamental causes that brought the conflict on in the first place may still be operating, the prevailing social mood swings in favor of cessation of conflict at all

costs, and an uneasy truce gradually takes hold. Those people, like the generation of Charles the Wise, who directly experienced civil war become “immunized” against it, and while they are in charge, they keep things stable. The peaceful period lasts for a human generation—between twenty and thirty years. Eventually, however, the conflict-scarred generation dies off or retires and a new cohort arises, people who did not experience the horrors of civil war and are not immunized against it. If the long-term social forces that brought about the first outbreak of internal hostilities are still operating, the society will slide into a second civil war. As a result, periods of intense conflict tend to recur with a period of roughly two generations (forty to sixty years).

These swings in the social mood may be termed “generation cycles” because they involve alternating generations that are either prone to conflict or not. Another example of such social mood dynamics has been noted, for example, by Arthur M. Schlesinger Jr. (1986). Furthermore, generation cycles keep cropping up in other contexts. The birth rates in twentieth-century America oscillated with a period of approximately fifty years (Easterlin 1980, Macunovich 2002). Many economic indicators oscillate with roughly the same period, a phenomenon known as the Kondratieff cycle (Kondratieff 1984). The Kondratieff and Schlesinger cycles may be related to each other; at least, they often seem to oscillate in synchrony (Berry 1991, Alexander 2002). The Kondratieff wave may also be correlated with the war cycle (Goldstein 1988). Our understanding of Easterlin, Schlesinger, and Kondratieff cycles is very deficient, and many researchers doubt the reality of these dynamics. This is not the place to try to make sense of this vast and confused topic, and in the rest of the book we focus only on the dynamics of sociopolitical instability. Even that focus is forced on us by the need to understand why disintegrative phases in certain types of societies tend to have multiple peaks of sociopolitical instability.

Exogenous Forces

The standard demographic-structural model of section 1.2 focuses on endogenous forces representing internal feedbacks between such structural variables as population, social structure, and instability. Real-life social systems are also affected by many exogenous factors that are not an explicit part of the model. We have alluded to some of them in this section; here is a more systematic discussion of the important external forces (see also the discussion in Turchin and Hall 2003).

- Geopolitical environment. Strong and aggressive neighbors may take advantage of internal weakness of the state during the disintegrative phase of the cycle. Such predation may deepen the degree of societal collapse. In the worst case the state may be conquered and annexed

to a nearby empire (a very powerful expansionist empire, such as that of Chinggis Khan, may simply roll over the studied state and obviate its endogenous dynamics). Alternatively, the presence of weak neighbors may permit external conquests that could relieve population pressure in the metropole and provide an outlet for surplus elites, thus lengthening the integrative phase.

- Disease environment. Some pandemics, such as the Black Death, originate in distant parts of Eurasia and then spread over the whole continent. Such pandemics arose repeatedly within Eurasia (Turchin 2008). Their effect depends on the phase of population growth. For a population in the early stages of growth, the arrival of an epidemic could mean a minor interruption of the course of expansion. By contrast, a dense population is highly vulnerable to a pandemic, and a severe drop in population numbers could result in a longer and deeper cycle of disintegration.
- Social evolution. Of primary interest is the growth of agricultural technology that affects the carrying capacity of the environment. Significant increases in crop yields, by elevating the carrying capacity, will have the same effect on food prices and consumption levels as substantial population declines.
- Global climate. Its effects are similar to those of social evolution, inasmuch as long-term fluctuations in temperature and rainfall affect the productivity of crops and the carrying capacity. A society whose population is already pressing on its resources may be tipped into crisis by a significant worsening of the climate.
- In addition to the recurrent exogenous factors discussed above, we often need to take into account singular events, or historical accidents that may have significant long-term consequences. A good example of such a singular event is the discovery and colonization of the Americas by Western Europeans, which resulted in torrents of precious metals flowing into Europe starting in the sixteenth century. American silver acted as an amplifier that created a stronger and more inflationary growth cycle in the sixteenth and early seventeenth centuries. In other words, the “price revolution of the sixteenth century” had two causes, monetary and demographic (Fischer 1996:74).

1.4 Empirical Approaches

The main goal of this book is to determine how well the predictions of the demographic-structural theory map onto empirically observed patterns in the studied historical societies. The synthetic theory, described in section 1.2, has four fundamental variables: population numbers (in relation

to the carrying capacity), social structure (specifically, the numbers and consumption levels of the elites), state strength (typically measured by its fiscal health), and sociopolitical instability. These variables are fundamental in the sense that it is the reciprocal interactions among them that generate secular cycles (in the parlance of dynamical systems theory, these are the endogenous variables). In each empirical case study our aim is to collect data describing how each of these variables changed during the period of study.

Ideally, we wish to have time-series data—accurate measurements of a particular variable collected at regular time intervals (the ideal time step is one decade, but a human generation—twenty to thirty years—serves almost as well). This ideal is rarely approached in historical applications. First, there is usually a substantial degree of measurement noise. This is not a fatal problem, because we can use statistical methods to estimate how much useful information is contained in the data. Even the worst case, when we lack quantitative data and all that we can say is that a variable is increasing, decreasing, or staying roughly constant, can be quite useful as a test of model predictions.

Second, we may have reasonable quantitative measures, but only for a few irregularly spaced points in time. Again, such data can be quite informative, especially if they are supplemented with qualitative indications about the dynamics of change between the “anchor points.” Reconstructions by knowledgeable historians can be surprisingly accurate, as happened in the case of estimates of population dynamics in early modern England that were later confirmed by the formal population reconstruction methods. Incidentally, there are statistical methods for time-series analysis that can help us utilize data to their utmost, even when they are irregularly spaced, although we do not employ them in this book.

It is frequently the case that although we lack direct measurements of some variable, with a little ingenuity we can come up with another one that could serve as a proxy for the variable of interest. For example, climatologists made great strides in reconstructing past climate variations by studying such proxy variables as tree rings, varves in lake deposits, and isotope compositions of air bubbles trapped in ice.

A very useful source of information is archaeological records (e.g., Morris 2005). Certain kinds of archaeological data, such as estimated numbers of dwellings during different time periods, can be quite good indicators of population dynamics. The population history of Novgorod is revealed by the density of leather shoe remains in cultural layers (Nefedov 2002). Such archaeological data often cannot tell us what the absolute level of population was (in people per km²). But having quantitative data on the relative fluctuations of a variable is almost as good for testing theory. In fact, it is much better to have a time series on relative fluctuations than an excellent

absolute estimate limited to one point in time. The demographic-structural theory is about dynamics, that is, change with time, and it is impossible to measure change with a single time slice.

Many quantitative data sources are available for testing theories about historical dynamics, and few of them have been systematically exploited. For example, the intensity of public building, especially of temples or churches, shows remarkable fluctuations in time. This index may reflect the amount of resources at the disposal of the state, the elites, or both, depending on the specific arrangements prevailing in the society.

Another underutilized indicator is the temporal distribution of coin-hoard finds. In 1969, Michael Crawford suggested that there is a close correlation between concentrations of coin hoards and periods of internal war and disturbance in the Roman Republic (see Crawford 1993:162). Another study documented a similar pattern in the late Carolingian period (Armstrong 1998).

Proxy variables have to be used carefully, because they may not be perfectly correlated with the variable of main interest. Thus, we expect that the number of people per building or per room should vary with time. As a result, in order to estimate the total population within a certain area, the estimated number of rooms obtained with archaeological methods needs to be multiplied by the average number of people per room, which is usually unknown. Similarly, the number of coin hoards per decade is affected not only by instability but also by the degree of monetization of the economy and by how much time has passed since the period when the hoards were interred (the farther this is in the past, the more chances that the hoard would have been found before modern times).

Although proxy variables need to be treated carefully, it would be madness to completely ignore them, because they are often the best quantitative information that we have about historical dynamics. One way to make sense of the proxies is to build an explicit model of the various factors that may affect them, estimate the model parameters, and then “impute” the values of the variable of interest. Statistical methods for doing this have been developed and applied to many natural science problems. After all, even in physics we usually cannot measure directly a quantity of interest, such as temperature; we have to infer temperature by a proxy variable, such as the expansion of a small amount of mercury in a glass tube. In complex geophysical applications, such as locating underground oil, nothing can be measured directly but has to be estimated by building a complex model of the underground geological layers. We will not be attempting such exercises in this book, but it is certainly something that can be tried in the future.

There are many other endogenous variables in addition to the fundamental ones and their proxies. Endogenous variables are those variables

that are part of the various feedback loops postulated by the theory. Their dynamics are largely determined by other endogenous variables (but there also can be an element of noise), and they in turn influence how other endogenous variables change with time. Exogenous variables, by contrast, are those that affect the state of the dynamical system but are not themselves affected by the state of the system. An example of an endogenous variable is the real wage. According to Malthusian-Ricardian theory, the real wage is primarily determined by the population numbers in relation to the productive capacity. It can also be influenced by other variables. For example, intense internal war may disrupt grain production and drive up food prices, with a deleterious effect on real wages. Real wages in turn influence other variables, such as demographic rates, which then affect the rate of population change. The point is that endogenous variables as a set describe the various feedback loops that drive the complex dynamics of the social system. A number of such variables, and predictions of the demographic-structural theory on how they should change with cycle phases, are given in table 1.1. In the chapters that follow our goal is to document the dynamics of as many as possible of these variables.

TABLE 1.1

Empirical predictions of the demographic-structural theory

	<i>Integrative Secular Trends</i>		<i>Disintegrative Secular Trends</i>	
	<i>Expansion phase (growth)</i>	<i>Stagflation phase (compression)</i>	<i>Crisis phase (state breakdown)</i>	<i>Depression/intercycle</i>
<i>Fundamental variables</i>				
Population dynamics	Population increases from nadir; rate of growth accelerates	Population is high and continues to increase but rate of growth decelerates	Population declines from the peak; the rate of decline accelerates	Population is low; it either declines at a decelerating rate or stagnates; periods of increase possible but do not lead to sustained growth
Elite dynamics	Low to moderate numbers; decline in elite/commoner ratio; modest consumption levels	“Golden age”: increasing numbers; increased competition for elite positions; conspicuous consumption by some segments; appearance of counterelites	High numbers; factionalization and conflict; high corruption; high income inequality; impoverishment of service elites	Reduction in elite numbers as a result of civil war and downward mobility; collapse of elite consumption levels
State strength and collective solidarity	Increasing; social unity among the elites that may extend to commoners	High but declining	Collapse; social disintegration	Periodic attempts to restore state, followed by repeated breakdown
Sociopolitical instability	Instability decreases to a low point	Instability is low but increasing	Instability increases to its peak	Instability is high but begins declining
<i>Other endogenous variables</i>				
Number of rural settlements	Increases	Slow increase or stagnation	Decline; settlement abandonment	Lack of increase
Land, cultivated	Increase; assarting	Slow increase or stagnation	Decline; settlement abandonment	At a low equilibrium
Land, free	Initially abundant but decreasing	In short supply	Increasing	Abundant
Land to peasant ratio	High but declining	Low	Low, increasing	High
Land prices	Low, increasing	High	Falling	Low
Grain prices	Low	Increasing	High, very variable	Decreasing, variable
Real wages	High	Declining to the lowest point	Increasing, but with much variability	High, but variable; contingent
Rents	Low	High; high exploitation by the landowners	Declining, but with fluctuations	Low, but variable; contingent
Personal consumption; subsistence level	High; infrequent crop failure incidents have no lasting effect	Declining; poverty, misery, vagrancy	Subsistence crises	Contingent (depends on instability levels)
Grain reserves	High	Declining	Nonexistent	Variable
Urbanization	Low	Increasing, growth of cities	High	High but declining

TABLE 1.1 (continued)

	<i>Integrative Secular Trends</i>		<i>Disintegrative Secular Trends</i>	
	<i>Expansion phase (growth)</i>	<i>Stagflation phase (compression)</i>	<i>Crisis phase (state breakdown)</i>	<i>Depression/intercycle</i>
Artisans and handcrafts	Low	Increasing; landless peasants become artisans	High	Declining
Trade	At a low level, local trading networks	Increasing in volume and spatial scale	Declining, variable, interrupted by political unrest	Local; long-distance networks disrupted
Usury	Absent	Increasing peasant indebtedness	High	Declining
Large private land-ownership	Absent, low, or medium	Increasing	High concentration of land in the hands of few large landowners	Declining
Economic inequality	Low	Increasing	High	High but declining
Incidence of epidemics	Rare; population bounces right back	Increasing; postepidemic population increases sluggish	Often catastrophic; population does not make up losses	High but declining
Internal peace and order	Increasing; a golden age	High but gradually unraveling; increasing resistance to taxation	Crisis: peasant uprisings, urban uprisings, interelite conflicts, regional/nationalist rebellions	Recurrent civil war, political fragmentation; high susceptibility to external invasions
Incidence of coin hoards (an indicator of sociopolitical instability)	Declining to low levels, unless there is a catastrophic external invasion	Low, unless there is a catastrophic external invasion	Rapidly increasing to a peak owing to state breakdown and civil war	Peaks when state breakdown and civil war recurs
State finances	Increasing revenues and stable expenditures, leading to budgetary surpluses	Declining real revenues, increasing expenditures due to growth of the army and bureaucratic apparatus	State bankruptcy loss of control over the army and bureaucracy	Finances generally in poor state, but high variability and contingency
Taxes	Increasing	Stagnant or even declining in real terms; heavy tax burdens on the peasantry	Tax system in a state of crisis	Variable; periods of high taxes alternate with collapse of the tax system
Ideology	Positive, optimistic ideologies rule the day	Growth of social pessimism; criticism of powers-that-be; ideological and social conflicts	Popular movements for social justice and abolishment of debts, and for land redistribution	Pessimistic ideologies; the cult of death
State policy	Internal policy is non-interventionist, laissez faire; externally, increased interest in conquest	Increasing attempts at social reforms, construction of irrigation and other infrastructure; colonization of borderlands; external aggression for acquisition of new territories	Social reforms, sometimes leading to social revolutions	Retrenchment; weakening of the state often results in external invasion