

Long-term population cycles in human societies

Peter Turchin

Department of Ecology and Evolutionary Biology

University of Connecticut

Storrs, CT 06269, USA

e-mail: Peter.Turchin@UConn.edu

To be published in:

This Year in Ecology and Evolutionary Biology 2009,

edited by R. S. Ostfeld and W. H. Schlesinger

(Annals of the New York Academy of Sciences)

ABSTRACT

Human population dynamics are usually conceptualized as either boundless growth or growth to an equilibrium. The implicit assumption underlying these paradigms is that any feedback processes regulating population density, if they exist, operate on a fast time scale, and therefore we do not expect to observe population oscillations in human population numbers. This review asks, are population processes in historical and prehistorical human populations characterized by second-order feedback loops, that is, regulation involving lags? If yes, then the implications for forecasting future population change are obvious—what may appear as inexplicable, exogenously driven reverses in population trends may actually be a result of feedbacks operating with substantial time lags. My survey of a variety of historical and archaeological data indicates that slow oscillations in population numbers, with periods of roughly 2-3 centuries, are observed in a number of world regions and historical periods. Next, I discuss a potential explanation for this pattern, the demographic-structural theory. Finally, I discuss the implications of these results for global population forecasts.

INTRODUCTION

Long-term human population dynamics are often portrayed as an almost inexorable exponential growth. During the 1960s, it even appeared that global population was growing at a faster-than-exponential rate, leading to predictions of “Doomsday” that was to occur on Friday, 13 November, 2026 (Von Foerster et al. 1960, Berryman and Valenti 1994). During the 1990s, when a noticeable decline in global population growth rate took place (due largely to precipitous drops in birth rates of populous developing countries, primarily China and India), it became clear that erstwhile predictions of doom (Ehrlich 1968) were overly alarmist. In fact, with most European countries experiencing population declines (which is especially noticeable in Eastern European countries, but would be as pronounced in Western Europe, except for a masking effect of immigration), the discussion in popular press shifted to the opposite tack. Now predictions are equally dire, but instead of fearing population explosion, we must worry about how to support increasing numbers of retirees on a shrinking base of the working population. Some of the current numerical predictions are as extreme as previous doomsday warnings. For example, popular press in Russia routinely publishes predictions that population of that country will be halved by 2050.

Although many media reports have a sensationalist, and even hysterical quality, the main question, what will happen to populations of different countries, as well as the total population of the Earth, is very important. The numbers and structure of human populations have an enormous impact on the well-being of individuals, societies, and the Biosphere. Yet, I would argue, the majority of forecasts are based on models of human population dynamics that are fundamentally flawed. The simplest forecasting devices are deterministic extrapolations of current trends. These approaches may employ the exponential model, or even faster-than-exponential growth, as in the Doomsday model. A somewhat more sophisticated approach is to allow for change in demographic rates (birth, death, migration), but still to assume that these processes are driven by exogenous influences (which can be modeled either stochastically, or by assuming a deterministic trend). Note that these standard approaches to forecasting human population are essentially *zero-order* models, because they do not take into account potential feedbacks from population density to demographic rates. Zero-order dynamics are nonequilibrium; depending on parameters the population numbers either increase to infinity, or decline to zero (Turchin 2003a:37). Including density-dependence in population growth models (the canonical model in population ecology is the logistic) leads to first-order dynamical processes, characterized by a convergence to an equilibrium (often called the carrying capacity). Human demographers began seriously entertaining density-dependent models much later than population ecologists working with nonhuman animals (Lee 1987).

First-order feedbacks act on a fast time scale. For example, in a territorial mammal, as soon as population has increased to the point where all available territories are occupied, any surplus animals become non-territorial “floaters” with poor survival rates and zero reproductive prospects. Thus, as soon as population numbers reach the carrying capacity, determined by the total number of territories, population growth rate is reduced to zero without any time lag. Some regulatory processes, however, act on a slow time scale (these are second-order feedbacks). The paradigmatic example of a second-order process in animal ecology is the interaction between predators and prey. When a population of prey reaches a high density so that predator population begins increasing, there is no immediate effect on prey’s population growth rate. This happens because it takes time for the predator numbers to increase to the level where they begin affecting prey numbers. Furthermore, once there are many predators, and prey population has started collapsing, the predators continue to drive prey numbers down. Even though there are few prey, and most predators are starving, it takes time for predators to die out. As a result, second-order population feedbacks act with a substantial lag and tend to induce oscillations.

As I have argued in *Complex Population Dynamics* (Turchin 2003a), such second-order processes as interactions between predators and prey, hosts and parasitoids or parasites, and plants and herbivores (generally known as trophic interactions) are very important drivers of population fluctuations. The great majority of population cycles in nature are driven by trophic mechanisms (Turchin 2003a:384). In contrast, human demographers, as far as I know, do not consider second-order processes when modeling and forecasting human population dynamics. There has been some discussion of

demographic cycles, for example, oscillations in the population age structure with the period roughly equal to one human generation (c.25 years). Another kind of a cycle is characterized by an alternation of high fecundity–low fecundity generations with an overall period of roughly 50 years (Easterlin 1980, Wachter and Lee 1989). In population ecology such oscillations are often called generation cycles and first-order cycles, respectively (Turchin 2003a:25). As we shall see below, second-order cycles should be characterized by much longer periods.

This review asks, are population processes in historical and prehistorical human populations characterized by second-order feedback loops? If yes, then the implications for forecasting future population change are obvious—what may appear as inexplicable, exogenously driven reverses in population trends may actually be a result of feedbacks operating with substantial time lags. Furthermore, once we have understood the specific mechanisms driving second-order dynamics, we may learn to anticipate such changes in demographic regimes.

A SURVEY OF POPULATION TRAJECTORIES IN HISTORICAL AGRARIAN SOCIETIES

Even a casual look at population history of humans over the last several millennia suggests that global population growth was not quite as inexorably exponential as it is usually portrayed (Figure 1). Apparently, there were several periods of rapid growth, interspersed with intervals of slower growth. Furthermore, Figure 1 gives a global view of human population dynamics. Because population change may be asynchronous in different regions, we need to focus on smaller parts of the human species. For example, we could delineate human populations by using country- or province-sized areas.

Figure 1. Global population numbers over the last four millennia (McEvedy and Jones 1978).

In addition to selecting the appropriate spatial scale, we also need to decide on the temporal scale at which we should study human population dynamics. The fundamental time scale in population dynamics is the generation time. The reproductive age for human females extends from 15 to 45 years of age (Wood 1990). However, the curve of fecundability as a function of age is skewed, so that its mode (maximum fertility rate) is at roughly 20, while its mean (average age of female at birth) is around 25–30 years (Wrigley et al. 1997). There is a large amount of variation in these measures of generation length for humans, depending on biological (nutrition, mortality schedules) and social (marriage age) characteristics of the population. It is clear, though, that for most historical human populations, generation time should lie in the interval between 20 and 30 years. This is, therefore, the time step at which, ideally, population trajectories should be measured.

Another, and longer, time scale is the expected period of second-order cycles, if such exist. Population theory based on discrete models suggests that second-order cycles are characterized by periods ranging from 6 to 12–15 generations (there is, actually, no upper limit on the period, but longer periods become progressively more rare in the

parameter space). Using the extreme values of 20 and 30 years for human generation time, suggests that second-order cycles in humans should have periods ranging from 120 to 450 years, with the most likely region of 200–300 years. Another way to evaluate the magnitude of period for possible second-order cycles, this time using the continuous framework, is to employ the formula suggested by the Lotka-Volterra predation model. The period of cycles in the Lotka-Volterra model near the neutrally stable point is $T = 2\pi/\sqrt{rd}$, where r is the intrinsic rate of population increase of prey, and d is the rate of decline of predator population in the absence of prey. Assuming that these two dynamical processes operate on the same temporal scale, $r \approx d$, we have $T \approx 2\pi/r$. In human populations r should be around $0.02\text{--}0.03 \text{ y}^{-1}$ (although in modern times some populations have been known to grow at 4% annual rate, and even higher, this is clearly an overestimate for preindustrial times, when infant mortality was much more severe). Thus, the estimate of cycle period is again 200–300 years. This is a very crude estimate, and by using the Lotka-Volterra model I do not mean to imply that population cycles in humans are driven by some mythical predators. Rather, I am using Lotka-Volterra equations as a generic model of second-order oscillations (it is the simplest possible system of ordinary differential equations that generates cycles). To sum, we expect that second-order cycles in human populations, if they exist, should be characterized by periods of two-three centuries. Thus, in order to detect such cycles we need multi-century time series. I now turn to data.

Western Europe

The first place to look for data is the population atlas of McEvedy and Jones (1978). The temporal resolution used in the atlas (a time step of 100 years after 1000 CE and 50 years after 1500 CE) is not sufficiently fine to analyze the data statistically, but for some areas where long-term population history is reasonably well known, like Western Europe, the overall pattern is quite striking (Figure 2). The figure plots population trajectories in just two countries, but others exhibit a qualitatively similar pattern. First, there is an overall increase in average population density. Second, around this millennial trend there are two secular cycles, with peaks c. 1300 and 1600.

Figure 2. Population dynamics in Czechoslovakia and Spain, 1100–1800 (McEvedy and Jones 1978).

The millennial trend reflects the gradual social evolution (which greatly accelerates after the end of the agrarian period, but in this review I focus on pre-industrial societies). Elsewhere (see Figure 3.10 in Turchin and Nefedov 2008) I made some quantitative estimates of this accumulation of knowledge and technology in the context of one case-study, England between 1150 and 1800. The secular oscillations appear to have the right period for second-order cycles, but we need more information before drawing any definitive conclusions.

China

Is this pattern of secular oscillations around a millennial trend specific to Europe, or does it hold more generally for agrarian societies? To answer this question, we travel to the opposite end of Eurasia. Ever since the first imperial unification by the Qin dynasty

in 221 BCE, the central authorities conducted detailed censuses for tax purposes. As a result, we have a record of Chinese population dynamics extending over two thousand years, although it has significant gaps during the periods of internal disunity and civil war.

Interpretation of the transmitted numbers is plagued by several difficulties. During the later stages of dynastic cycles corrupt or lazy officials often falsified or even fabricated outright population data (Ho 1959). Conversion coefficients between the number of taxable households and the actual population are often unknown, and it is possible that these coefficients changed from dynasty to dynasty. The area controlled by the state also continually changed. Finally, it is often difficult to determine whether the number of taxable households declined during the times of trouble as a result of demographic change (death, emigration), or as a result of the state's failure to control and enumerate the subject population. Thus, there is a certain degree of controversy among the experts as to precisely what the numbers mean (Ho 1959, Durand 1960, Song et al. 1985). The disagreements, however, primarily concern absolute population levels, while there is a substantial degree of agreement on the *relative* changes in population density (which are, of course, of primary interests to our purposes). Chinese population, essentially, expanded during periods of political stability and declined (sometimes precipitously) during periods of unrest. As a result, population movements closely mirror the “dynastic cycle” in China (Ho 1959, Reinhard et al. 1968, Chu and Lee 1994).

The most detailed history of Chinese population, known to me, has been published by Zhao and Xie (1988). When examined over the whole two-thousand year period, the population trajectory is clearly nonstationary. In particular, there were two abrupt changes in population regime (Turchin 2003c:159). Before the eleventh century, population peaks were in the vicinity of 50–60 mln (Figure 3a). In the twelfth century, however, population peaks double to around 100–120 mln (Turchin 2003c: Figure 8.3). The mechanism underlying this regime change is known. Prior to the eleventh century, the center of gravity of the Chinese population was situated in the North, with the South lightly settled. Under the Sung dynasty the South has matched, and then overtaken the North (Reinhard et al. 1968: Figures 14 and 115). Additionally, new, high-yielding varieties of rice were developed during the Sung dynasty. The second regime change occurred in the eighteenth century, when population started growing at a very rapid pace, reaching 400 mln during the nineteenth, and more than 1 billion during the twentieth century. In order not to deal with these regime changes, here I focus on the quasi-stationary period from the start of the Western Han dynasty until the end of the Tang, 201 BCE – 960 CE (for the following centuries, see Turchin 2003c: Section 8.3.1).

Figure 3. (a) Population history of China between 201 BCE and 960 CE (Zhao and Xie 1988). The dynastic cycles are indicated above each population peak. The dotted line indicates the estimated numbers during the period of disunity between the collapse of the Eastern Han and the Sui unification, when population reconstruction is very uncertain. (b) Instability index of China between 200 BCE and 1000 CE (Lee 1931).

During these twelve centuries the Chinese population reached at least four peaks, each in the region of 50–60 million people (Figure 3a). Each population peak was achieved during the later phases of the great unifying dynasties, the Western and Eastern Han, the Sui, and the Tang. Between the peaks the population of China collapsed to less than 20 million (although some scholars dispute these trough estimates, for reasons given above). While the quantitative details of the Zhao and Xie reconstruction are disputed, the qualitative pattern—long term population oscillations associated with the dynastic cycle, involving drastic changes in population numbers—is not in doubt.

Northern Vietnam

We have seen, thus, that secular cycles in population numbers show up at both ends of the great Eurasian land mass. Another example of the same pattern was noted by Victor Lieberman in his book suggestively titled *Strange Parallels: Southeast Asia in Global Context, c.800-1830* (Lieberman 2003). The pattern of population oscillations in Northern Vietnam (Figure 4) is in many ways reminiscent of that observed in Western Europe (Figure 2): there is an ascending millennial trend and secular oscillations around it. However, unlike the two oscillations that were observed in Western Europe, there were three in Vietnam (Figure 4).

Figure 4. Population dynamics in Northern Vietnam, 1100–1850 (Lieberman 2003).

PROXY DATA ON POPULATION DYNAMICS FROM ARCHAEOLOGY

Population reconstructions, such as those shown in Figures 1–4, have one significant disadvantage: they are plagued by a variety of subjective biases. To make such reconstructions, the authorities typically have to integrate a great variety of very heterogeneous sources of information, some quantitative and some qualitative. The basis for giving different weights to different kinds of data are not always made explicit. As a result, we often have quite different trajectories proposed by different authorities. This does not mean that we should dismiss informed judgements of highly knowledgeable experts. In fact, the population dynamics of England during the early modern period (sixteenth–eighteenth centuries) reconstructed informally by experts turned to be very close to the trajectory later estimated by the formal methods of family reconstruction (Wrigley et al. 1997). Nevertheless, it would be useful to have some other, more objective ways to track population dynamics of historic (and prehistoric) human societies.

Archaeological data offers the basis for such alternative methods. People leave a variety of traces that can be measured. Thus, the basic idea of the approach is to focus on *proxy data*, variables that may be directly correlated with past population numbers. Typically such approaches yield relative estimates of population dynamics—by what percentage the population changed from one period to another—rather than absolute estimates of people per squared kilometer. This is quite sufficient for the purposes of my review, since I am interested precisely in relative fluctuations of numbers. In some cases, furthermore, it is possible to estimate the coefficients of proportionality and obtain absolute estimates.

Site occupation dynamics in western Roman Empire

One serious problem that often makes archaeological data less useful is the coarse temporal resolution. For example, a reconstruction of population history of the Deh Luran Plain in western Iran (Dewar 1991) suggests at least three huge oscillations in density (10-fold peak/trough ratios). However, the sampling interval is around 200–300 years, much too coarse for our purposes.

Fortunately there are archaeological case-studies in which the sampling rate is much finer (and it is to be hoped that with time there will be more such examples). The first case study relates to the question of population history of the Roman Empire, the subject of an intense scholarly debate (Scheidel 2001). Tamara Lewit surveyed both published and “grey literature” reports of archaeological excavations of settlements within the western part of the Roman Empire, and calculated what proportion of these settlements was actually occupied for each time period (first century BCE, first century CE, and then for 50-year periods until the fifth century). It turns out that the occupation index went through two grand oscillations during these five centuries (Figure 5a).

Figure 5. (a) Dynamics of the occupation index in Western Roman Empire (Lewit 1991). Occupation index is defined as the proportion of the total number of excavated settlements that was occupied during each period from first century BCE to fifth century CE. (b) Importation of African Red Slip Ware into the Albegna Valley (Etruria) (Bintliff and Sbonias 1999:5).

Lewit (1991) also presented the data broken down by region within the western Roman Empire (Britain, Belgica, Northern and Southern Gaul, Italy, and Northern and Southern Spain). The qualitative dynamics were similar (two secular cycles) but quantitatively the peak and trough levels varied by region. Significantly, much of this variation closely corresponded with what we know of the history of these regions. For example, the degree of the third century decline correlated with how much the region was affected by barbarian invasions as documented in the historical chronicles.

In summary, Lewit’s compilation suggests that the population of the Roman Empire (at the very least, its western half) went through two secular cycles between the first century BCE and fifth century CE. However, the occupation index is not a very good proxy for population numbers, because the relationship between these two variables may be nonlinear (for example, it is possible that during the population trough, it was the smaller settlements that were more likely to be abandoned; in such a case, the occupation index would overestimate the degree of change from peak to trough). Additionally, the time resolution, one century to half-century, is still rather coarse.

As archaeological methods get better, time resolution of data improves. For example, the fluctuations in the rate of importation into Albegna Valley (in modern Tuscany) of a kind of pottery, called African Red Slip Ware, has been resolved down to the decadal time step (Figure 5b). The pottery importation trajectory shows a similar pattern to the occupation index, with the peaks in the late second and late fourth century. Unfortunately, trade is not a good proxy for population numbers, because maritime traffic

between North Africa (where the pottery was produced) and Italy (where it was imported) could have been disrupted by the political crises of the third century.

Novgorod the Great

An example of archaeological data that is both well resolved and appears to be a good proxy for population numbers comes from the medieval city of Novgorod the Great in northwestern Russia (Yanin 1990). Novgorod is located in a very wet and cool environment. In the middle of the tenth century the Novgorodians became tired of constantly wading through the mud, and paved the streets of their city with huge wooden planks. This worked for a while, but the soil level slowly rose because of accumulating anthropogenic rubbish, and some twenty years later the streets again became muddy. The Novgorodians then put another layer of street paving on top of the old one, and continued doing so at intervals of 20–30 years throughout the next six centuries. Because of the cool and humid environment all layers of wooden pavements were perfectly preserved. Twentieth century's archaeologists dated each layer using dendrochronological methods. As a result we have an unusually well resolved stratigraphy for Novgorod with the time step of roughly one human generation.

If we look at the rates with which various kinds of “stuff” accumulated in Novgorod's soil, we observe two clear peaks, one during the twelfth century and the second around 1400. The decline during the fifteenth century is real, but after 1500 is an artefact. Unfortunately for archaeologists, a drainage system was installed in Novgorod in the early modern period, as a result of which the post-1500 cultural layers became aerated and all organic matter in them decomposed.

Figure 6. The rate of deposition of various kinds of anthropogenic rubbish in Novgorod: leather shoe remnants, broken locks, discarded pieces of cloth, and lost amber ornaments (mainly beads). Each data series was normalized by scaling to mean = 1 in order to fit on the same graph. Data sources: (Izyumova 1959, Rybina 1978).

It is reasonable to suggest that the greater the population density in the city, the more rubbish they would generate. If people wear out shoes at an approximately constant rate, the more people there were during a certain period of time, the more discarded shoes should archaeologists find in the corresponding cultural layer. The general (although not perfect) agreement between the trajectories of the four kinds of stuff shown in Figure 6 suggests that there is a real signal indicating population changes. Furthermore, there is a variety of other qualitative and some quantitative data that support the interpretation that Novgorod population went through two secular cycles between 950 and 1500 (Nefedov 2002).

THEORETICAL EXPLANATIONS OF SECULAR CYCLES

The general conclusion from the survey of historical and archaeological data is that secular cycles in population numbers are observed in a wide variety of world regions and historical periods. It appears that we have discovered a macrohistorical regularity, which raises the question, can we propose a mechanistic theory explaining this recurrent dynamical pattern?

The Malthusian theory

The obvious place to start is with the ideas of Thomas Robert Malthus (1798). The basic argument goes like this. Population growth beyond the means of subsistence results in increasing food price and the decline of real wages (that is, wages expressed in terms of real commodities, such as kg of grain), so that per capita consumption decreases, especially among the poorer strata. 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 a decline) that, in turn, allows subsistence 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 result 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). In the twentieth century the argument was taken up by such Neomalthusians as Wilhelm Abel, Michael Postan, and Emmanuel Le Roy Ladurie (Postan 1966, Le Roy Ladurie 1974, Abel 1980).

There are problems with this argument, some empirical (which will be dealt with in a later section) and some theoretical. The theoretical problem becomes apparent when we rephrase the Malthusian argument using the terms of modern population dynamics. Let us assume that the rate of technological progress occurs on a slower time scale compared to the scale of secular cycles (this appears to be a reasonable assumption for preindustrial societies). Thus, the carrying capacity K will be set by the amount of land available for agriculture and the current agrarian technology (yields per unit of area). As population approaches the carrying capacity, all available land will be cultivated. Further increase of population numbers immediately (without a time lag) results in lower average consumption rates. Since there is no time lag, the population numbers will not overshoot the carrying capacity, instead approaching it gradually. In other words, we are dealing here with a first-order dynamical process, the simplest model for which would be the logistic equation, and the end result of our assumptions are not cycles, but a stable equilibrium. A discrete first-order process can generate cycles of alternating high-density/low-density generations, but such a model is inappropriate for human population with overlapping generations. Furthermore, as we have seen in the data survey, secular cycles are characterized by ascending and descending phases several generations long. These are clearly second-order processes, and thus the explanatory model should be a system of equations with population numbers N interacting with some other dynamical factor X (or X, Y, \dots if there are more than one). The Malthusian theory does not provide a clear answer just what this factor X might be.

The Demographic-Structural Theory

Although Malthus mentioned war as one of the consequences of population growth, he did not develop this line of reasoning in any detail. The Neomalthusian theory of the twentieth century was concerned almost exclusively with the demographic and economic variables. A significant extension of the Malthusian model was achieved by the historical sociologist Jack Goldstone (1991), who focused on indirect effects of

population growth on society's structures. More specifically, Goldstone argued that excessive population growth has several effects on social institutions. First, it leads to persistent price inflation, falling real wages, rural misery, urban migration, and increased frequency of food riots and wage protests (this is essentially the Malthusian component). Second and even more important, rapid expansion of population results in an increased number of aspirants for elite positions. Increased intraelite competition leads to the formation of rival patronage networks vying for state rewards. As a result, elites become riven by increasing rivalry and factionalism. Third, population growth leads to expansion of the army and the bureaucracy and rising real costs. States have no choice but to seek to expand taxation, despite resistance from the elites and the general populace. Yet, attempts to increase revenues cannot offset the spiraling state expenses. Thus, even if the state succeeds in raising taxes, it is still headed for fiscal crisis. 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 that follow the breakdown of central authority (Goldstone 1991).

Goldstone's focus was on the effects of population growth on sociopolitical instability. But it stands to reason that there is also a feedback effect of instability on population dynamics (Turchin 2003c). Most obviously, when the state is weak or absent, the population will suffer from elevated mortality due to increased crime, banditry, and internal and external warfare. Additionally, the times of troubles cause increased migration rate, as refugees flee war-affected areas. Migration may lead to emigration (and we can simply add that to mortality) and to spread of epidemics. Increased vagrancy spreads the disease by connecting areas that would stay isolated during better times. As vagabonds and beggars aggregate in towns and cities, they may tip the population density over the epidemiological threshold (a critical density above which a disease spreads). Finally, political instability causes lower reproduction rates, because during uncertain times people choose to marry later and to have fewer children. People's choices about their family sizes may be reflected not only in birth rates, but also in increased rates of infanticide.

Instability can also affect the productive capacity of the society. First, the state offers protection. In a stateless society people can live only in natural strongholds, or places that can be made defensible. Examples include hillfort chiefdoms in Peru before Inca conquest (Earle 1991), and the movement of settlements to hilltops in Italy after the collapse of the Roman Empire (Wickham 1981). Fearful of attack, peasants can cultivate only a small proportion of productive area that is near fortified settlements. A 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. Second, states often invest in increasing the agricultural productivity by constructing irrigation canals, roads, and flood control structures. A protracted period of civil war results in a deterioration and outright destruction of this productivity-enhancing infrastructure (Turchin 2003c).

In summary, the demographic-structural theory (so called because it asserts that the effect of population growth is filtered through social structures) models the society as a system of interacting parts that include general population, the elites, and the state

(Goldstone 1991, Nefedov 1999, Turchin 2003c). What are the implications of these assumptions for the dynamical behavior of the system? The only way to answer questions of this type is by constructing mathematical models. In Chapter 7 of my book on *Historical Dynamics* (Turchin 2003c) I developed a suite of dynamical models addressing various aspects of the demographic-structural theory. Here I discuss two models that were developed in a recent paper (Turchin and Korotayev 2006).

A model of population dynamics and internal warfare in agrarian empires

The first model is an extension of the mathematical theory of state collapse discussed in Turchin (2003c: Chapter 7). There are three structural variables: population numbers, state strength (measured by the amount of resources taxed by the state), and the intensity of internal warfare (that is, such kinds of sociopolitical instability as major outbreaks of brigandage, peasant uprising, regional rebellions, and civil war).

Let $N(t)$ be the number of inhabitants at time t , $S(t)$ be the accumulated state resources (which we can measure in some real terms, e.g., tons of grain), and $W(t)$ the intensity of internal warfare (measured, for example, by extra mortality resulting from this type of conflict). To start deriving the equations we assume that the per capita rate of surplus production, ρ , is a declining function of N (this is Ricardo's law of diminishing returns) (Ricardo 1817). Assuming, for simplicity, a linear relationship, we have

$$\rho(N) = c_1(1 - N/K)$$

Here c_1 is some proportionality constant, and K is the population size at which surplus equals zero. Thus, for $N > K$, the surplus is negative (the population produces less food than is needed to sustain it). To derive the equation for N we start with the exponential form (Turchin 2003a):

$$dN/dt = rN$$

and then modify it by assuming that the per capita rate of population increase is a linear function of the per capita rate of surplus production, $r = c_2\rho(N)$. Putting together these two assumptions, we arrive at the logistic model of population growth:

$$dN/dt = r_0N(1 - N/K) \tag{1}$$

where $r_0 = c_1c_2$ is the “intrinsic rate” of population growth, and parameter K is now seen to be the “carrying capacity” (Gotelli 1995).

State resources, S , change as a result of two opposite processes: revenues and expenditures. If the state collects a fixed proportion of surplus production as taxes, then revenues equal $c_3\rho(N)N$, where $\rho(N)N$ is the total surplus production (per capita rate multiplied by population numbers), and c_3 the proportion of surplus collected as taxes. State expenditures are assumed to be proportional to the population size. The reason for this assumption is that as population grows, the state must spend more resources on the army, police, bureaucracy, and public works. Putting together these processes we have

$$dS/dt = \rho_0N(1 - N/K) - \beta N \tag{2}$$

where $\rho_0 = c_1c_3$ is the per capita taxation rate at low population density and β the per capita state expenditure rate.

The dynamics of internal warfare intensity, W , in the absence of state is a balance of conflict initiation and termination rates. First, we assume that conflict initiation rate is

proportional to the square of population density. This functional form is derived by analogy with chemical kinetics (and also with the predation term in the Lotka-Volterra model). The rate at which each individual “bumps” into others, which may potentially lead to a conflict between them, is proportional to N ; and the total rate of bumping is proportional to N^2 . This is a very crude analogy, and clearly causes of human conflict are much more complex than that, but it serves as a useful starting point for modeling (we also investigated other functional forms; the qualitative results were the same).

Second, we assume that the intensity of warfare, in the absence of hostility initiation events, declines gradually at the exponential rate b . This assumption reflects the “inertial” nature of warfare. Even in the absence of new hostility initiation events, warfare is perpetuated by chains of revenge and counter-revenge. Time is needed for the memory of past wrongs to fade so that the participants become willing to “bury the hatchet”.

Third, the presence of the state should have a restraining effect on the intensity of internal war. We model this process by assuming that W declines at the rate proportional to state resources, S . Putting these three assumptions together we have the following equation for W :

$$\frac{dW}{dt} = aN^2 - bW - \alpha S \quad (3)$$

The final ingredient of the model is the feedback loop from W to N . We model the effects of W on both the demography and on the productive capacity of the society. The demographic effect is modeled by assuming an extra mortality term proportional to W . Additionally, we assume that K , the carrying capacity, is negatively affected by warfare: $K(W) = k_{\max} - cW$. Putting together all these assumptions we have the following equations:

$$\begin{aligned} \frac{dN}{dt} &= rN \left(1 - \frac{N}{k_{\max} - cW} \right) - \delta NW \\ \frac{dS}{dt} &= \rho N \left(1 - \frac{N}{k_{\max} - cW} \right) - \beta N \\ \frac{dW}{dt} &= aN^2 - bW - \alpha S \end{aligned} \quad (4)$$

All variables are constrained to nonnegative values.

Dynamics of Model (4) are illustrated in Figure 7. At the beginning of a cycle, both population (N) and state resources (S) grow. Increasing S suppresses internal warfare, with the result that the carrying capacity $K(W)$ increases to the upper limit k_{\max} . At the same time, extra mortality due to warfare declines to zero. As a result, N increases rapidly. However, at a certain population size, well before N approaches k_{\max} , the growth of state resources ceases (because the state revenues fall behind its expenditures), and S begins to collapse at an increasing rate, rapidly reaching 0. State collapse allows internal warfare to grow, and it rapidly reaches its maximum level. This means that $K(W)$

decreases, and mortality increases, leading to a population density collapse. Population decline eventually allows S to increase again, S suppresses W and another cycle ensues.

Figure 7. Dynamics of Model (4). The trajectory is shown for the period between 500 and 1000 years to allow for transient dynamics to die out. Parameter values: $\rho = 1$, $\beta = 0.25$, $r = 0.015$, $k_{\max} = 3$, $\delta = 0.1$, $a = 0.01$, $b = 0.05$, and $\alpha = 0.1$.

Depending on parameter values, model dynamics are characterized by either a stable equilibrium (with oscillatory approach) or stable limit cycles, such as shown in Figure 7. The main parameter affecting the cycle period is the intrinsic rate of population increase. For realistic values of r between 1 and 2 percent per annum we obtain cycles with a period of roughly two hundred years. In other words, the model predicts a typical pattern of second-order oscillations with average period of cycles similar to those observed in historical data.

A model of internal warfare in stateless societies

The second model investigated the dynamics of population and internal warfare in stateless societies. The model has two structural variables: N , population density, and W , warfare intensity. To construct the equation for N we first assume that in the absence of war population will grow logistically. Second, death rate due to warfare is assumed to be directly proportional to warfare frequency. By appropriately scaling W we can redefine it as the annual warfare death rate, leading to the following equation:

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right) - WN \quad (5)$$

The dynamics of W are modeled as above, except there is no state to put a damper on warfare:

$$\frac{dW}{dt} = aN^2 - bW \quad (6)$$

Mathematical analysis of the model indicates that its dynamics are characterized by a single equilibrium that is stable for all values of parameters. However, the approach to the equilibrium is oscillatory. Thus, in the presence of exogenous variables that stochastically perturb the trajectory from the stable point, we should observe noisy-looking cycles in which peaks of N are followed, with a lag, by peaks in W . In other words, the model for stateless societies also exhibits second-order oscillations in population numbers and warfare intensity, although the cyclic pattern is not as pronounced as in the model for state societies.

EMPIRICAL TESTS OF THE THEORY

The models discussed in the previous section (as well as those in Chapter 7 of Turchin 2003c) suggest that demographic-structural mechanisms can generate second-order cycles of the right period. The models do more: they make specific quantitative predictions that can be tested with historical data. One striking pattern, predicted by theory, is that sociopolitical instability should oscillate with the same period as population density but shifted in phase, so that the peaks of instability follow population peaks.

To test these predictions empirically we need case studies for which both population and instability dynamics are known. However, as was discussed earlier, population reconstructions for most historical societies are plagued by various problems that make their quantitative details suspect. Furthermore, demographic-structural processes is not the only mechanism that affects the dynamics of historical societies. As a result, we often observe much more complex patterns than regular cycles predicted by the models involving variability in oscillation periods and phase shifts between the structural variables. Given the limitations of historical data and the complexity of the dynamical pattern, we need to employ an appropriately coarse-grained procedure for testing theoretical predictions. One approach works like this. First, we identify the population growth and decline phases. Although quantitative details of population dynamics for historic societies are rarely known with any precision, there is usually a consensus among demographic historians about when the qualitative pattern of growth changed. Second, we count instability events (peasant uprisings, separatist rebellions, civil wars, etc) that occurred during each phase. The instability data are taken from such compilations as that of Sorokin (1937), Tilly (1992), or Stearns (2001). Finally, we compare the incidence of instability events per decade between the two phases. The demographic-structural theory predicts that instability should be higher during the population decline phases.

This procedure was applied to all seven complete cycles examined in Turchin and Nefedov (2008) (see Table 1). The empirical regularity is very strong: in all cases instability is greater during the declining, compared to growth, phases (t-test: $P \ll 0.001$).

Table 1. Instability events per decade during the growth and decline secular phases (after Table 10.2 in Turchin and Nefedov 2008).

<i>Secular Cycle</i>	<i>Growth phase</i>		<i>Decline phase</i>	
	<i>years</i>	<i>Instability</i>	<i>years</i>	<i>Instability</i>
Plantagenet	1151–1315	0.78	1316–1485	2.53
Tudor	1486–1640	0.47	1641–1730	2.44
Capetian	1216–1315	0.80	1316–1450	3.26
Valois	1451–1570	0.75	1571–1660	6.67
Republican	350–130 BCE	0.41	130–30 BCE	4.40
Principate	30 BCE–165	0.61	165–285	3.83
Muscovite	1465-1565	0.60	1565-1615	3.80
Average (\pmSE)		0.6 (\pm0.06)		3.8 (\pm0.5)

Using the same procedure we can also test for association between population oscillations and the dynamics of political instability during the imperial period of Chinese history (from the Han to the Qing Dynasties). Population data are taken from Zhao and Xie (1988), the instability data from Lee (1931). The test is conducted only for periods when China was unified under one dynasty (Table 2).

Table 2. Instability events per decade during the growth and decline secular phases

	<i>Growth phase</i>	<i>Decline phase</i>
--	---------------------	----------------------

<i>Secular Cycle</i>	<i>years</i>	<i>Instability</i>	<i>years</i>	<i>Instability</i>
Western Han	200 BCE–10	1.5	10–40	10.8
Eastern Han	40–180	1.6	180–220	13.4
Sui	550–610	5.1	610–630	10.5
Tang	630–750	1.1	750–770	7.6
Northern Sung	960–1120	3.7	1120–1160	10.6
Yuan	1250–1350	6.7	1350–1410	13.5
Ming	1410–1620	2.8	1620–1650	13.1
Qing	1650–1850	5.0	1850–1880	10.8
Average (\pmSE)		3.4		11.3

Again, we observe a remarkable match between the prediction and what is observed: the level of instability is invariably greater during the population decline compared to increase phases.

Note that the secular phases in this empirical test were defined as periods of growth and decline, that is whether the first derivative of population density is positive or negative. The response variable in the test, is not a derivative but the level of instability. This means that instability should peak roughly in the middle of the population decline phase. In other words, instability peaks are shifted with respect to population peaks that take place, of course, when population growth phase ends and decline begins.

The importance of the phase shift is that it provide clues to the potential mechanisms that drive oscillations. If two dynamical variables oscillate with the same period, and their peaks are not lagged, that is, happen at roughly the same time, then such a situation is *inconsistent* with the hypothesis that it is the dynamic interaction between the two variables that drives the observed oscillation (Turchin 2003b). On the other hand, if the peak of one variable is lagged with respect to the other, than this is consistent with the hypothesis that the interaction between these two variable drives the cycle. A classical example in ecology is the predator-prey cycles modeled by Lotka-Volterra and similar models, in which predator peaks follow prey peaks (Turchin 2003a: Chapter 4).

The demographic-structural models discussed above also show this dynamical pattern. Note, for example, the phase shift between population (N) instability (W) in Figure 7. In fact, instability in that model is only positive during the population decline phase.

In situations when we have higher quality data it is possible to go beyond the coarse-grained procedure. For example, the demographic-structural theory predicts that there should be dynamical feedbacks between the two variables: increasing sociopolitical instability has a negative effect on the rate of population growth, while increasing population density has a positive effect on the growth rate of instability. Thus, we can attempt to detect these relationships by regressing the rates of change of variables on their levels. In a recent paper (Turchin 2005) I used this approach in the analysis of four time-series datasets: early modern England, the Han and Tang China, and the Roman Empire.

I fitted the data with regression models like this one:

$$\Delta X(t) = a_0 + a_1 X(t) + a_2 Y(t) + \varepsilon_t$$

where $X(t)$ is log-transformed population density, $Y(t)$ is log-transformed index of instability, $\Delta X(t) = X(t+1) - X(t)$ is the rate of change of $X(t)$, and ε_t is the error term. The analysis results indicated that including instability in the model for the population rate of change greatly increased the degree of fit, and vice versa, population density was a statistically significant predictor for the rate of change of instability. In other words, these results provide further evidence in favor of the mechanisms postulated by the demographic-structural theory.

CONCLUSIONS

The main conclusion I draw from the empirical review is that the typical pattern found in historical populations is neither exponential growth, nor small fluctuations around some stable carrying capacity. Instead, we usually observe long-term oscillations (around a gradually rising level). These “secular cycles” (called so because they consist of alternating increase and decline population phases, each roughly a century long) appear to be the rule for agrarian state-level societies, and we see them whenever we have reasonably good quantitative data on population dynamics. Where such data are absent, we can infer the presence of secular cycles from the empirical observation that the great majority of agrarian states in history were affected by recurring instability waves (Turchin and Nefedov 2008).

Secular oscillations are not strict, mathematically precise cycles. Rather, they appear to be characterized by an average period with fairly large amount of variation around it. This is to be expected because human societies are complex dynamical systems with many parts cross-linked by nonlinear feedbacks. It is well known that such dynamical systems are prone to mathematical chaos, or in more technical terms, sensitive dependence on initial conditions (Ruelle 1989). Furthermore, social systems are open in the sense that they are influenced by exogenous influences such as climatic change, or sudden appearances of evolutionary novel pathogens. Finally, humans have free will, and their actions and decisions at the individual micro-level can have societal macro-level consequences. Sensitive dependence (chaos), exogenous influences, and human free will all combine to produce very complex dynamics whose future trajectories are very difficult (and probably impossible) to predict with any degree of accuracy. Additionally, there is the well-known problems of self-fulfilling and self-defeating prophecies—situations when making a forecast affects the trajectory that is being forecasted.

Although future trajectories of human social systems (including its demographic part, which is the focus of this article) may be difficult to predict with any accuracy, it does not mean that these dynamics are not worth studying. The strong empirical patterns that I reviewed in this article suggest that there may be general principles operating, that history is not “just one damn thing after another.” If so, then understanding such governing principles can inform governments and societies about the consequences of social choices that they contemplate. There is no reason to believe that there is any inevitability about the social dynamics that I have discussed in this article. Of particular

interest are such undesirable consequences of sustained population growth as instability waves.

Political instability in collapsed or collapsing states is one of the greatest sources of human misery today. Since the end of the Cold War, interstate warfare accounted for less than 10 percent of conflicts. The great majority of wars today are within-state conflicts, such as civil wars and separatist rebellions (Harbom and Wallensteen 2007). I do not see why the humanity must continue to endure these periods of state collapse and civil strife for ever. However, at this point we still know too little about social mechanisms leading to instability waves. We lack good theories to tell us how to rebuild states and avoid civil wars, although there are promises of one (Turchin 2008). This is an area where research effort has a good chance of paying off, both in generating empirically tested theoretical breakthroughs, and in alleviating human misery.

Returning to the issue of long-term forecasts of the global population, probably the most important conclusion that follows from my review is this. The smooth logistic-like curves produced by various government and U.N. agencies (and copied in many ecological textbooks), in which the global population nicely equilibrates around some number, 10 or 12 billion, are essentially worthless as serious forecasts. The global population is a dynamical quantity that is determined by the balance of mortality and natality rates. There is no reason to believe that these two rates will equilibrate and cancel each other out. In the last two crises experienced by the global population, in the seventeenth and the fourteenth centuries, global population declined substantially, and in many regions precipitously. In the fourteenth century many regions of Eurasia lost between a third and a half of their populations (McNeill 1976). In the seventeenth century, fewer places in Eurasia were hit as bad (although Germany and central China lost between a third and a half of their populations). The population of North America, on the other hand, was reduced perhaps ten-fold, although this remains a matter of controversy. Thus, a forecast based on historical patterns suggests that the twenty-first century should also be a period of population decline.

On the other hand, probably the most important aspect of recent human history is the rapid acceleration of social evolution in the last two centuries that usually goes under the name of industrialization (or modernization). The global carrying capacity (Cohen 1995) during this period has increased dramatically, and it is very difficult to predict how it will continue to change. Thus, it is conceivable that the trend to greater carrying capacity will continue and that it will overwhelm the negative delayed effects of twentieth century's population growth. We do not know which of these two opposing tendencies will prevail, but what is clear is that they will not simply cancel each other out. Thus, equilibration of the global population around some constant level in the twenty-first century is actually a most unlikely outcome.

References cited

- Abel, W. 1980. *Agricultural fluctuations in Europe: from the thirteenth to the twentieth centuries*. St. Martin's Press, New York.
- Berryman, A. A., and M. A. Valenti. 1994. The doomsday prediction. *Bulletin of the Ecological Society of America* **75**:123-124.
- Bintliff, J., and K. Sbonias, editors. 1999. *Reconstructing past population trends in Mediterranean Europe (3000 BC - AD 1800)*. Oxbow, Oxford.
- Chu, C. Y. C., and R. D. Lee. 1994. Famine, revolt, and dynastic cycle: population dynamics in historic China. *Journal of Population Economics* **7**:351-378.
- Cohen, J. E. 1995. *How many people can the Earth support?* W. W. Norton, New York.
- Dewar, R. E. 1991. Incorporating variation in occupation span into settlement-pattern analysis. *American Antiquity* **56**:604-620.
- Durand, J. D. 1960. The population statistics of China, AD 2-1953. *Population Studies* **13**:209-256.
- Earle, T., editor. 1991. *Chieftdoms: power, economy, and ideology*. Cambridge University Press, Cambridge.
- Easterlin, R. 1980. *Birth and Fortune*. Basic Books, New York.
- Ehrlich, P. R. 1968. *The Population Bomb*. Ballantine Books.
- Goldstone, J. A. 1991. *Revolution and Rebellion in the Early Modern World*. University of California Press, Berkeley, CA.
- Gotelli, N. J. 1995. *A primer of ecology*. Sinauer, Sunderland, MA.
- Harbom, L., and P. Wallensteen. 2007. Armed Conflict, 1989—2006. *Journal of Peace Research* **44**:623-634.
- Ho, P. T. 1959. *Studies on the Population of China: 1368-1953*. Harvard University Press, Cambridge, MA.
- Izyumova, S. A. 1959. On the history of leather and shoe manufacture in Novgorod the Great (in Russian). *Materials and Studies on the Archaeology of the USSR* **65**:197.
- Le Roy Ladurie, E. 1974. *The Peasants of Languedoc*. University of Illinois Press, Urbana.
- Lee, J. S. 1931. The periodic recurrence of intrnecline wars in China. *The China Journal* (**1931: March-April**):111-163.
- Lee, R. D. 1987. Population dynamics of humans and other animals. *Demography* **24**:443-465.
- Lewit, T. 1991. *Agricultural productivity in the Roman economy A.D. 200-400*. Tempus Reparatum, Oxford, UK.
- Lieberman, V. 2003. *Strange Parallels: Southeast Asia in Global Context, c.800-1830*. Cambridge University Press, Cambridge.
- Malthus, T. R. 1798. *An essay on the principle of population*. J. Johnson, London.
- McEvedy, C., and R. Jones. 1978. *Atlas of World Population History*. Facts on File, New York.
- McNeill, W. H. 1976. *Plagues and peoples*. Anchor Books, New York.
- Nefedov, S. 1999. *The method of demographic cycles in a study of socioeconomic history of preindustrial society*. PhD dissertaion, Ekaterinburg University (in Russian), Ekaterinburg, Russia.

- Nefedov, S. A. 2002. On the demographic cycles in the history of the medieval Russia (in Russian). *Klio* **3**:193-203.
- Postan, M. M. 1966. Medieval agrarian society in its prime: England. *in* M. M. Postan, editor. *Cambridge Economic History*. Vol. I. Cambridge University Press, Cambridge.
- Reinhard, M. R., A. Armengaud, and J. Dupaquier. 1968. *Histoire Générale de la Population Mondiale*. Editions Montchrestien, Paris.
- Ricardo, D. 1817. *On the principles of political economy and taxation*. John Murray, London.
- Ruelle, D. 1989. *Chaotic Evolution and Strange Attractors*. Cambridge University Press, Cambridge, UK.
- Rybina, E. A. 1978. *Arkheologicheskie ocherki istorii novgorodskoy trgovli X-XIV vv.* Moscow University Press, Moscow.
- Scheidel, W., editor. 2001. *Debating Roman demography*. Brill, Leiden, Netherlands.
- Song, J., C. H. Tuan, and J. Y. Yu. 1985. *Population Control in China: Theory and Applications*. Praeger Scientific, New York.
- Turchin, P. 2003a. *Complex Population Dynamics: A Theoretical/Empirical Synthesis*. Princeton University Press, Princeton, NJ.
- Turchin, P. 2003b. Evolution in population dynamics. *Nature* **424**:257-258.
- Turchin, P. 2003c. *Historical dynamics: why states rise and fall*. Princeton University Press, Princeton, NJ.
- Turchin, P. 2005. Dynamical feedbacks between population growth and sociopolitical instability in agrarian states. *Structure and Dynamics* **1(1)**:Article 3.
- Turchin, P. 2008. Building nations after conflict. *Nature* **453**:986-987.
- Turchin, P., and A. Korotayev. 2006. Population Dynamics and Internal Warfare: a Reconsideration. *Social Science and History* **5(2)**:121-158.
- Turchin, P., and S. Nefedov. 2008. *Secular cycles*. Princeton University Press, Princeton, NJ.
- Von Foerster, H., P. M. Mora, and L. W. Amiot. 1960. Doomsday: Friday, 13 November, AD 2026. *Science* **132**:1291-1295.
- Wachter, K. W., and R. D. Lee. 1989. US births and limit cycle models. *Demography* **26**:99-115.
- Wickham, C. 1981. *Early medieval Italy: Central power and local society, 400-1000*. Macmillan, London.
- Wood, J. W. 1990. Fertility in anthropological populations. *Annual Review of Anthropology* **19**:211-242.
- Wrigley, E. A., R. S. Davis, J. E. Oeppen, and R. S. Schofield. 1997. *English population history from family reconstruction: 1580-1837*. Cambridge University Press, Cambridge, UK.
- Yanin, V. L. 1990. The Archaeology of Novgorod. *Scientific American* **262**:84.
- Zhao, W., and S. Z. Xie. 1988. *Zhongguo ren kou shi: China Population History (in Chinese)*. People's Publisher, Peking.

Figure 1

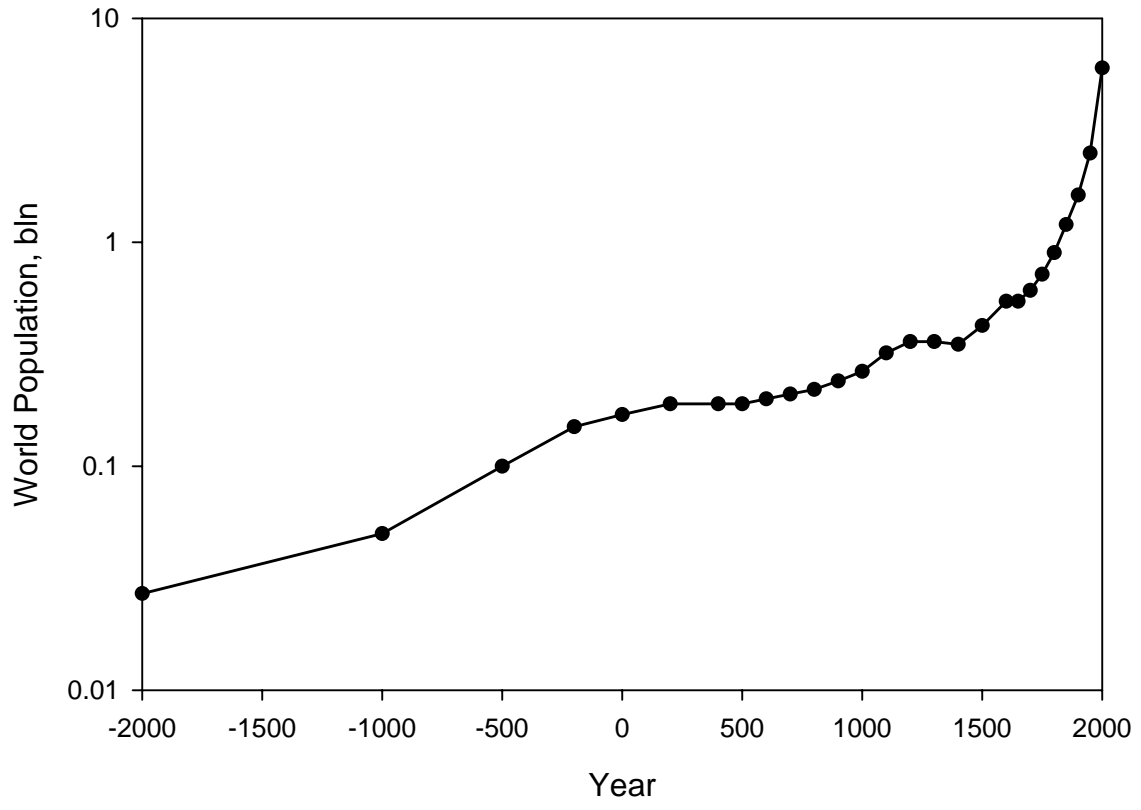
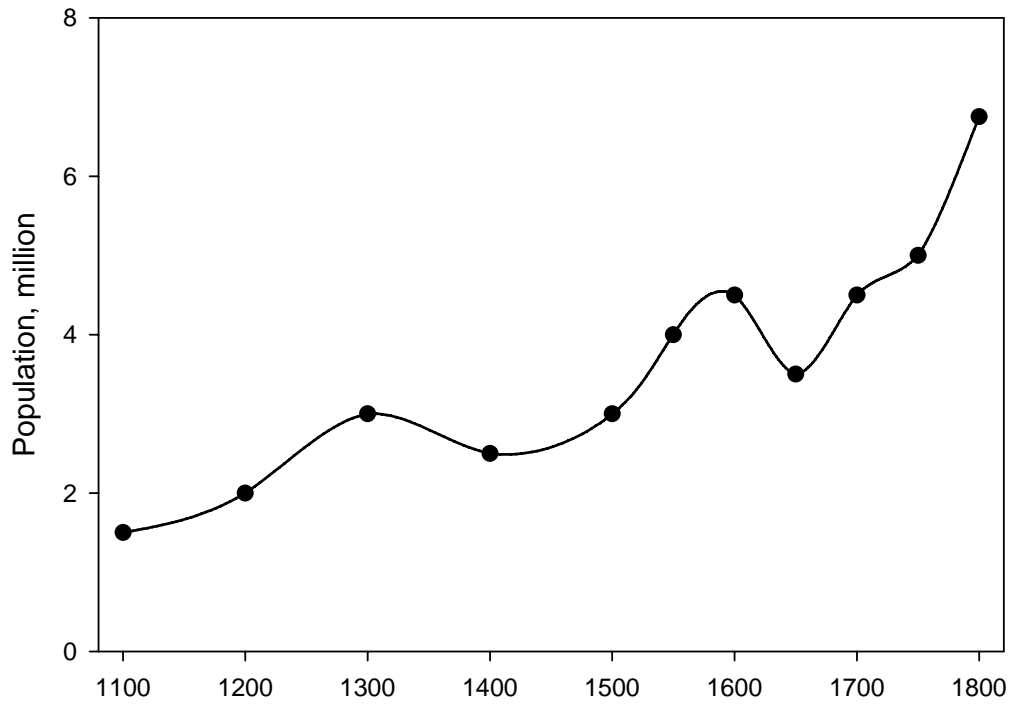


Figure 2

Czechoslovakia (McEvedy and Jones 1978)



Spain (McEvedy and Jones 1978)

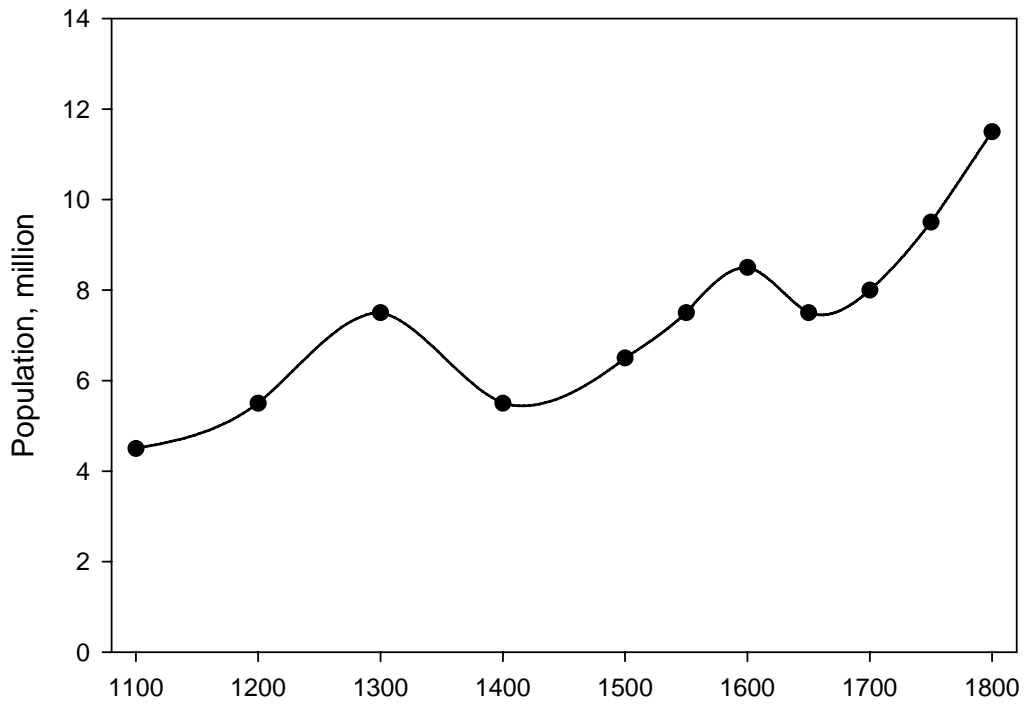


Figure 3

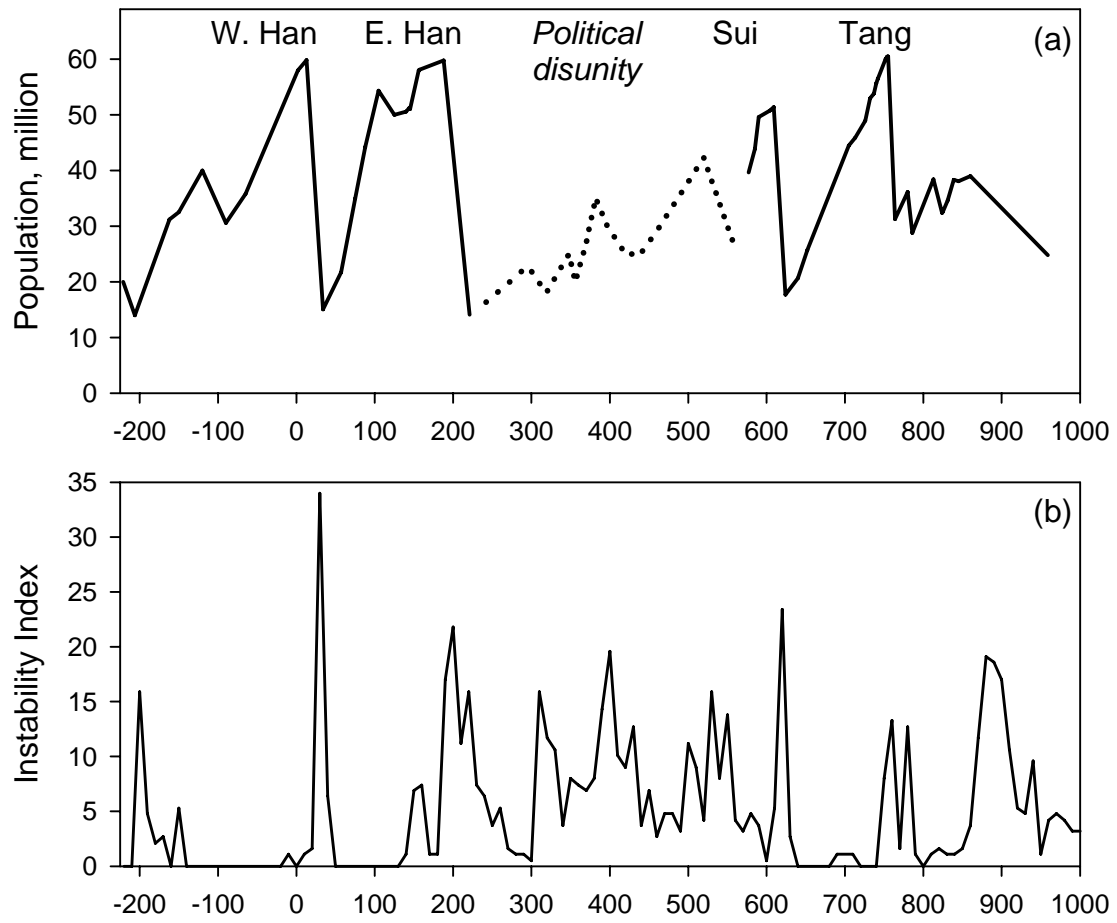


Figure 4

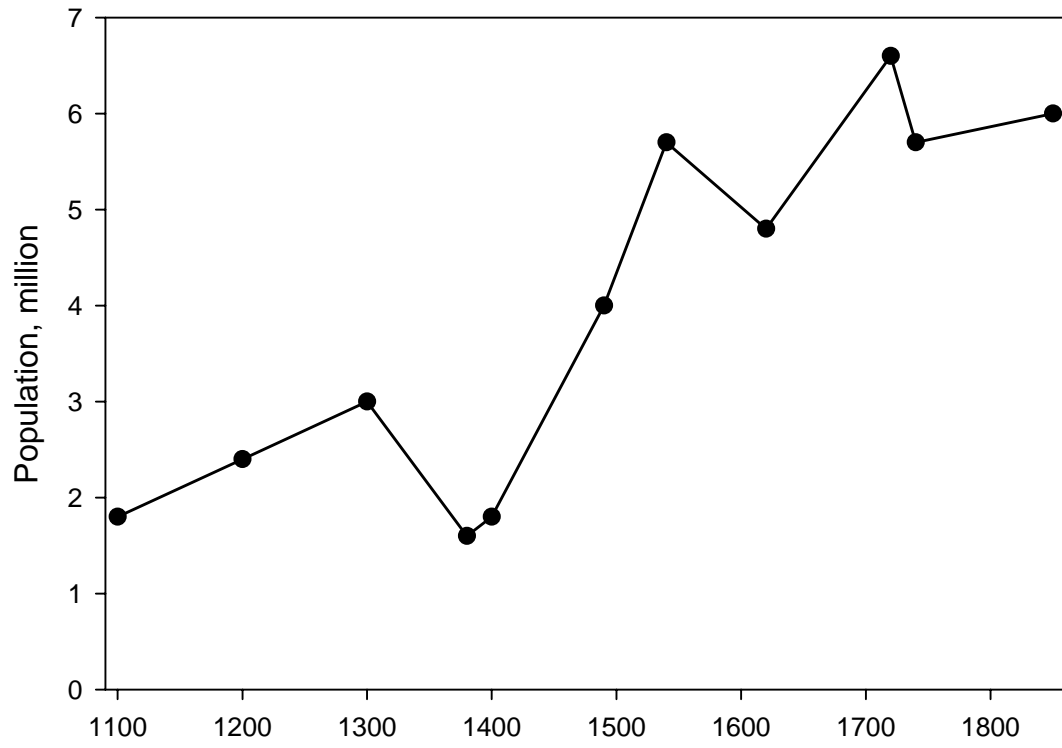


Figure 5

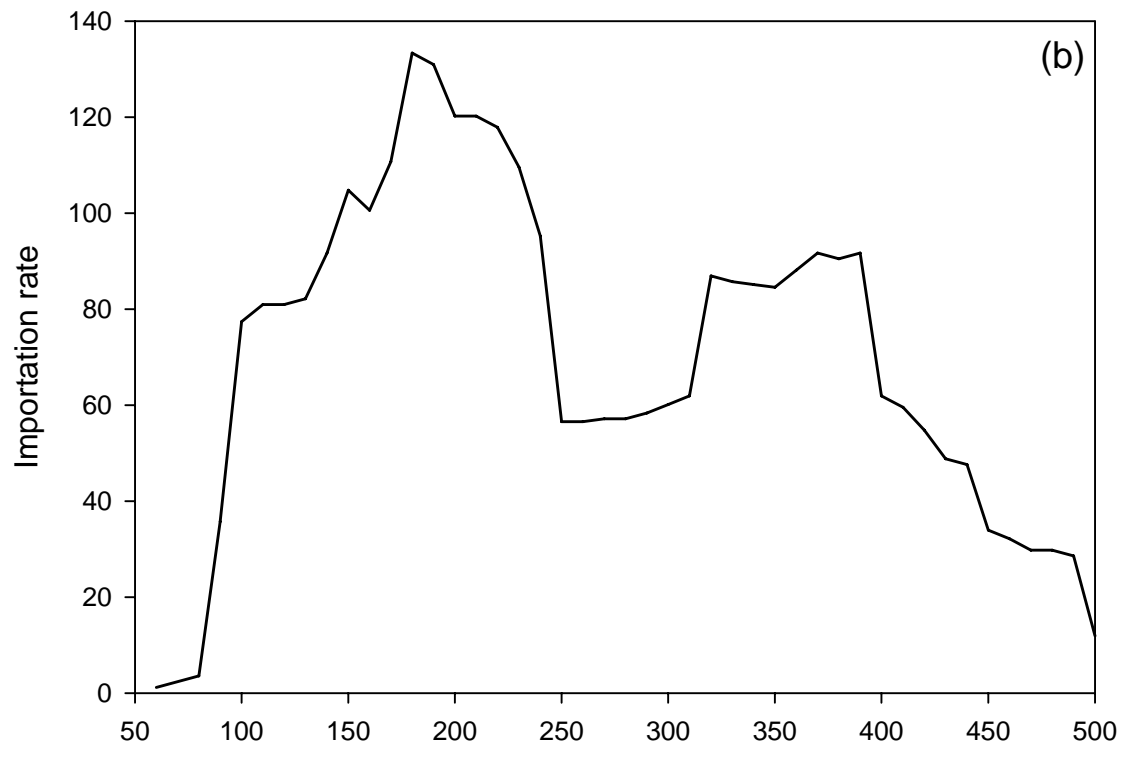
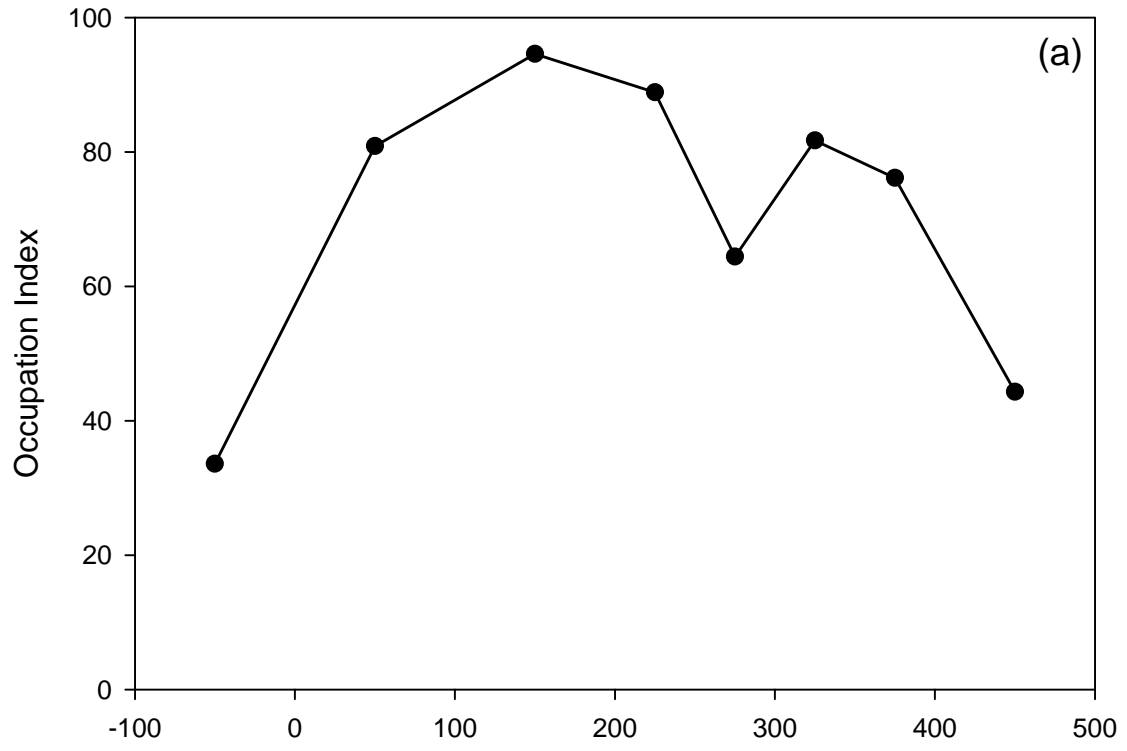


Figure 6

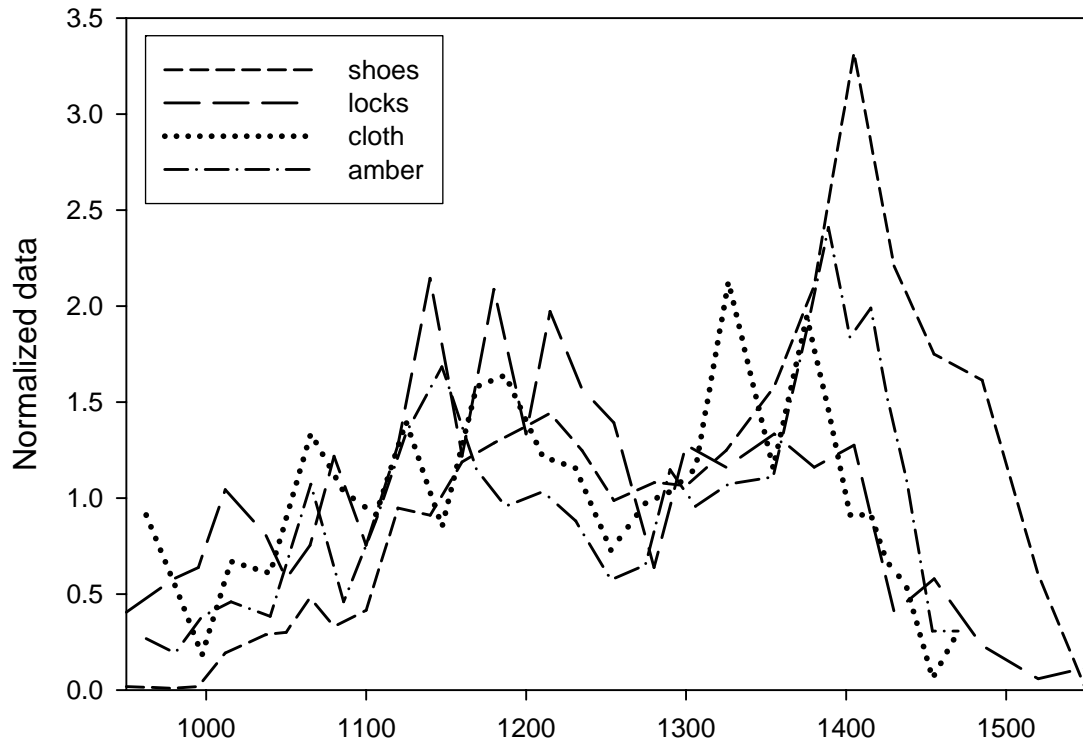


Figure 7

