1. Time and space—an introduction

The theory of capital—including the value of capital and its accumulation—depends on a number of fundamental concepts. The *durability* of a good is the time during which the good contributes capital to production. All goods are to some extent durable and thus all goods have capital values. The durability of a good is the inverse of the depreciation rate of the good (in a deterministic world; for the stochastic case see Lev and Theil, 1978). All material goods are physical and thus deteriorate according to the laws of entropy; all such goods will eventually break down, even with high spending on maintenance. The choice of how durable a good should be therefore has an upper limit, and laws of nature set this limit. Buyers and sellers of goods can however choose a shorter than maximum durability for economic reasons. These economic reasons include the discount rate as well as the valuation of durability, with the latter being associated with various attributes of the good. We discuss these matters in more detail in Chapter 5.

The choice of how *durable* a production process should be is a key economic problem. Among other things, the duration of the production process includes the time spent preparing for the later stages of the process such as research and development, prototype production, marketing and sales promotion. A classic type of problem is the choice of how long to store cheese or wine. Another well-known problem concerns the decision when to harvest trees, grapes or fish. Among practitioners, it is quite common to aim for the point in time that corresponds to the maximum sustainable yield. In Chapter 2 we show that this decision rule is suboptimal, if the alternative real interest rate on capital is positive. The choice of the duration of a production process is more general than the cases associated with storage or harvesting. In complex processes, such as those that are typical of the pharmaceutical, automotive and film industries, there are unavoidable additional stages of production including numerous research, development and other intermediate sub-processes that producers must complete before a good reaches the market. At each stage, there is an optional decision of whether to proceed further or to abort the production process (cf. Dixit and Pindyck, 1995). Both the durability of a good and the duration of its production process have considerable impacts on the growth rates of production and capital as well as on the location of economic activities. We focus on this topic in Chapter 6.

Expectations and *uncertainty* connect the past to the future, and reflect the level and heterogeneity of knowledge among the relevant decision makers. The valuation of capital and the decision to invest in new capital reflect investors' uncertain expectations. It has been shown that even mild increases in uncertainty as approximated by estimated risk levels of capital portfolios can cause substantial decreases in capital values as determined in the stock markets as well as cause less willingness to invest in new capital, leading to reduced economic growth.

Profits, dividends and expected returns are the entities that determine the market valuation of capital. The classical theory of growth assumes that savings out of profits is the main driver of the growth of both capital and income. The value of capital is then proportional to the ratio of the profit and discount rates.

Dynamic interactions between scientists, inventors and entrepreneurs affect the growth of capital. New technological or design ideas spread most easily among spatially proximate firms and can lead to clustering phenomena such as those associated with computer software in Silicon Valley, fashion design in Milan and feature films in Los Angeles and Mumbai.

Surprisingly, mainstream economic theory not only lacks these dynamic concepts; it also even avoids the use of any explicit account of processes of change in time and space. General equilibrium theory—as elaborated upon by Gérard Debreu and other mathematical economists in the 1950s and 1960s—includes time and location in a superficial way as variable subscripts, but there are no *mechanisms* that represent processes of change that involve temporal or spatial factors. General equilibrium theory only addresses the dynamic aspects of an economic system in studies of the stability of equilibrium solutions to static models.

It is however possible to represent time as an explicit dimension of economic processes in many different ways in economic theories and models. The first and most obvious way is to represent changes over time, such as the accumulation of capital and other dynamic economic processes, as *continuous*, mostly real-valued, variables. This implies the use of differential equations in the modeling of economic processes. One early example is Walrasian *tatônnement* (Walras, 1874/1896), in which the process of changing prices at a point in time is a function of the excess demands for the goods at the *same* point in time, with the added assumption that no exchanges will take place before the attainment of equilibrium. The price changes will cease when all excess demands approach zero, which assumes asymptotically stable dynamic bidding processes.

Other neoclassical economists discussed and modeled growing populations, capital and income, using more or less sophisticated differential equations. Cassel (1918/1932) proposes a linear model, which Domar (1946) develops as a simplified ordinary differential equation. In this type of linear differential equation model, the product of the savings rate and the productivity of capital determine the unbounded rate of growth. Leontief (1953) presents a dynamic multi-good model of economic growth that generalizes the results from Domar's linear growth model.

Neoclassical growth theory—as exposed by Tinbergen (1942), Swan (1956) and Solow (1957)—is an example of a dynamic process that would put an upper limit to the growth rate given an exogenously determined savings rate. In these models, the per-capita aggregate product is a concave function of the (somehow) aggregated per-capita capital. These theorists in fact never discussed or motivated their aggregation of capital. In these models, the accumulation of capital per unit of labor equals the product of the savings rate and net aggregate production per person. The implication is that equilibrium with no further per-capita income growth corresponds to the point of equality between the savings rate and the rate of depreciation of capital per person, as Figure 1.1 illustrates.

If the depreciation rate goes down—that is, if capital durability goes up—the capital intensity, k* and output per worker, y*, will both increase. If durability approaches infinity, then capital intensity and output per worker will also go toward infinity—a most surprising and quite incredible idea.

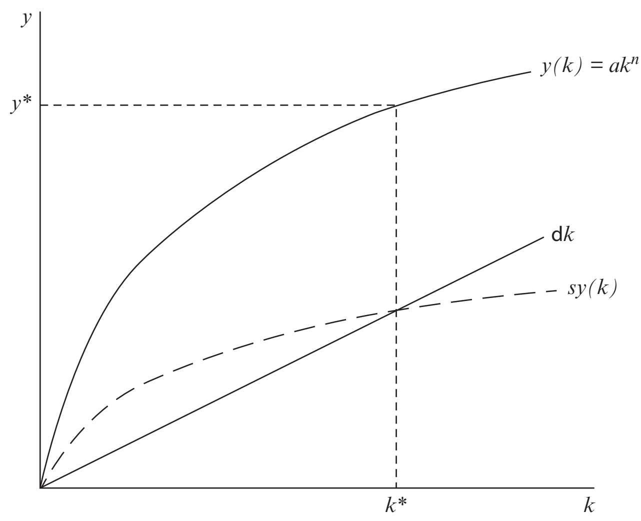

Figure 1.1 Equilibrium in a neoclassical growth model with a Cobb-Douglas production function

According to this model, a growth policy based on an increased savings rate is futile. It is only growing knowledge that can result in per-capita income gains. This conclusion makes no economic sense. Knowledge is a form of capital and its accumulation requires savings, in the same way that investments in physical capital requires savings. This fact implies that a realistic growth theory must include endogenously accumulated knowledge as a type of capital.

EXPLICIT DISCRETE TIME IN ECONOMIC MODELS

The second way to represent dynamic economic processes is as a discrete set of periods (for example as weeks, quarters or years). In this case, difference equations are useful. Such equations are quite popular in economics for two reasons. First, most statistical measurements use an accumulated flow over a period such as a year or quarter. Second, economists often formulate their hypothesized causal structures using sentences of the "if-then" type, which they interpret as "if some economic event happens in this period, then economic effects occur one, two or more periods later." The econometrician Herman Wold first formally stated and defended this type of period-based causality; it was later renamed "Granger causality."

In an article in *Econometrica*, Wold (1954) puts forward his causality beliefs and preference for difference equations. He starts out with a model in which the current price determines the demand for some good, while the price one period earlier determines the supply. Consequently, the price in the current period equals the price in the previous period plus a constant times the excess demand one period earlier. Hanau (1928) is the first instance of this type of recursive relation between the supply and demand for a good. Hanau's article analyzes the price fluctuations of *Schweinefleisch* (pork)—the "Hog Cycle"— as illustrated by Figure 1.2.

Wold modeled the Hog Cycle as difference equations, where demand in this period depends on the current price, while the price observed one period earlier determines the supply. This leads to a difference equation for the evolution of the market price. The price would converge toward equilibrium if the slope of the supply function were smaller than the slope of the demand function. From this example of a recursive process, Wold proceeded to his arguments for a causal analysis that would later form the basis of an econometric school:

> The general characteristics of models of this kind are as follows:
> (a) The model refers to a sequence of years, months, or other time units.

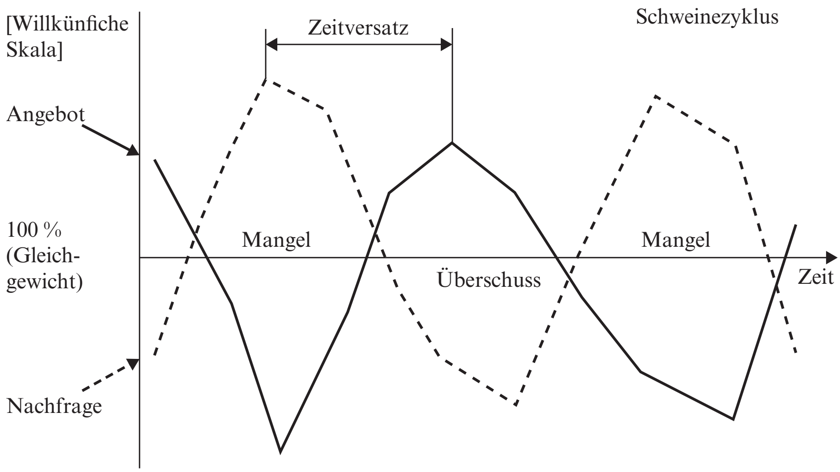

Figure 1.2 The Hog Cycle as seen by Hanau with a reproductive time delay between demand and supply. A positive difference between demand (Nachfrage) and supply (Angebot) leads to an increase in the price of pork

(b) All relations of the model are causal, with two types of variables, endogenous, which it is the purpose of the model to explain, and exogenous, which are auxiliary. In every relation of the model the effect variable is thus endogenous, while the cause variables are either endogenous or exogenous.
(c) The model has one, and only one, causal relation for each endogenous variable.
(d) Given the development of the exogenous variables and a set of initial values for the endogenous variables, the model allows us to calculate, recursively, the development of the endogenous variables (Wold, 1954, pp. 172–3).

But Wold's strongest reason for recursive difference equation models was probably that the parameters of such models allow the use of ordinary least squares estimation method without complications such as lack of identification.

A couple of decades later Granger (1969) offers a similar view of causality in econometrics. Granger defines the causality relationship with the help of two principles:

1 The cause happens prior to its effect.
2 The cause has *unique* information about the future values of its effect.

Granger (ibid.) claims that the ability of a model to predict the future values of one time series using past values of another demonstrates causality in econometrics. The terms *predictive causality* or *Granger causality* refer to this type of ability. Most philosophers would object to this type of causality concept, based as it is on time-recursive relations between variables as represented by discrete difference equations (we discuss this further in Chapter 3). Causal analysis of the Wold-Granger type is of course applicable to situations where interdependencies are very limited, as is the case in the experimental sciences, engineering and other fields where it is easy to avoid interference from unanalyzed variables. This kind of separation is almost never possible in economics. The complexity of an economic system with a multitude of sellers, buyers, goods and resources is the generic case, implying irregular fluctuations and even chaos. Paradoxically, most statistical economic indicators maintain a stability of their quantities over time that may seem to contradict the complexity of the economy. Why are these magnitudes, for example gross domestic products or exchange rates, as stable as they are in statistical terms?

A NEW WAY TOWARD PREDICTABILITY IN A SEEMINGLY CHAOTIC WORLD

A quite novel way to model dynamic processes in many of the sciences and more recently also in economics is to allow for many *different continuous and interactive timescales* of dynamic economic processes. We can use a macro model of the labor market as an example of how these processes play themselves out.

Most macroeconomic models, whether classical, monetary or Keynesian, make use of comparative static time assumptions within an implicitly delimited spatial area. For example, there is no role for explicit time in the Keynesian multiplier analysis, which reflects a recursive sequence of dependence reactions. Even in the latest versions of common macroeconomic models, there are “short-term” or “medium-term” solutions that depend on a hypothesized approach toward “natural levels” of unemployment or inflation, among other factors.

In neo-Keynesian analyses of the labor market the natural level of unemployment depends on the difference of inflation and production from their respective natural rates, as observed from the perspective of a nation state, be it the United States, Britain, Belgium or even Iceland. The assumption is that the same processes are operating, irrespective of differences in spatial and temporal dimensions.

The focus of neo-Keynesian macroeconomic analysis is usually on the equilibrium of unemployment, inflation and real wage rates. In monetarist and new classical analyses, the treatment of time and space is similar, and thus implicit, but such models assume that nominal wage rates and the prices of goods are flexible enough to ensure an equilibrium level of employment, at least in the (undefined) medium term. A natural rate of unemployment is then a consequence of movements in and out of the labor force and of the short time spans needed when workers search for better job opportunities. Unemployment thus only occurs because of necessary market frictions.

In contrast, a Keynesian analysis assumes rigid wages and prices with the consequence that unemployment becomes an equilibrating variable. A high level of unemployment is then possible as a permanent equilibrium condition.

It is against this theoretical backdrop that the political left-right conflict on suitable labor market policies developed, first in the 1930s and then again after the financial crisis that affected Europe and North America for several years after 2008.

Both dominant approaches lack an understanding of economic change processes. A more dynamic theoretical approach to the unemployment problem is possible and necessary in order to explain contemporary unemployment in many different regions around the world. This more realistic approach to the analysis of national labor markets focuses on *parallel processes of change* that both occur on different timescales and affect *spatially separated* labor markets in different ways.

Relative wages, prices of goods and the hiring and firing of non-specialists are quite flexible during the different phases of the business cycle. They are examples of *fast process variables*. Capital of various types is however quite rigid during the timescale of a business cycle and are examples of *slow process variables*. Contemporary science and especially the rapidly expanding field of synergetics show that an understanding of the dynamics of most interactive systems requires a division of timescales into fast and slow processes of change. It further requires a distinction of system variables between private (individual) and public (collective) ones. Variables that are both slow and public jointly make up the infrastructure of the system. Haken (1983; 1993) and Mikhailov and Calenbuhr (2002) provide important and informative examples of synergetic theory as applied to processes of change in physics, chemistry, the life sciences and social theory.

The basic idea of synergetics is that a variable that is slowly changing and public—implying that it influences many of the private and/or fast variables—in a generic sense shapes the structural outcome for the system as a whole. In our context, this implies that the set of slowly changing public economic variables determine the underlying equilibrium structure of all regional labor markets (see Table 1.1).

Table 1.1 Fast and slow processes with private and public effects in the labor market

Effects/speed	Fast	Slow
Private	Market prices and wages; Market supply of and demand for labor; Unemployment rate; Marginal return to capital	Market supply of and demand for human and physical capital; long-run returns to capital
Public	Labor market information; General economic and political information; Exchange rates	**Accessibility to markets and to knowledge; Formal and informal institutions**

The upper left corner of the table contains the short-term equilibrating processes of each regional labor market. Given a specific set of slow private and public variable values, the regional labor market process will move toward a short-term equilibrium that reflects whether—and to what extent—the slow and public institutional variables make wage and price flexibility possible. If the labor market institutions make it difficult to adjust wages there is a strong case for understanding the underlying equilibrium as a high-unemployment Keynesian one. On the other hand, if these institutions promote quick price adjustments in all markets, the corresponding employment equilibrium would resemble the one typically associated with monetarist and new classical theories (see Appendix 7.1 for a more formal exposition of these ideas).

Accessibility within various spatial networks that convey people, goods and information change at a very slow rate in most places, most of the time. The total yearly increment to the transport infrastructure is in most countries well below .001 in relation to the total network length; the impact on accessibility is even smaller because of natural geographical conditions. Massive infrastructure investment programs, such as the new network of highways and railways that is being put in place in China at the time of writing, may for a short period of time increase this fraction somewhat, but this is—and always has been—a transitional phenomenon. On the other hand, many regions have remained relatively inaccessible for thousands of years and are therefore disadvantaged in their relative access to national and international markets for goods and services.

Spatial differences in accessibility to scientific, artistic and other public knowledge have consequences for the diffusion of ideas, for innovative activities and for firms' productivity. There are good infrastructural economic reasons that explain why highly accessible European regions such as London, Paris, Amsterdam and Stuttgart have persistently higher levels of human capital productivity than more peripheral regions.

When discussing the flexibility or rigidity on wages it is important not to lose sight of the institutions that govern relations among employers and workers in the labor market. Nevertheless, these institutions are only one aspect of a subset of a much wider array of institutions that jointly make up the institutional structure of a group, region or nation. Such institutions may be formal or informal. Sometimes they are both, as when theft of property is both illegal and deemed unethical by most people, regardless of the minutiae of various legal sanctions. Formal institutions refer to codified legal and political systems as well as various laws and regulations. Informal institutions reflect cultural norms, habits and values, and entail various types of informal sanctions that constrain human behavior (North, 1990). Shared public values shape the evolution of both formal and informal institutions, but they only *become* institutions when they prescribe or proscribe specific actions among a group of people.

Political scientists and sociologists (see, for example, Inglehart, 1997; Inglehart and Welzel, 2005) have shown that there is a great deal of interregional variability in the public values that affect institutions, and that these values change at a slow pace because of the dominance of cohort rather than life cycle effects. Thus, shared values are arguably also part of the set of slowly changing and public variables that jointly make up the infrastructure. We should therefore be unsurprised about the attempts to make culture accountable for differences in economic outcomes between nations or cultures. Prominent examples include Max Weber's celebration of Protestant or Calvinist culture and his derision of the Confucian mindset (Weber, 1920/1958) and more recent attempts to view Confucianism as being supportive of—rather than inimical to—economic development (for example Morishima, 1982). As the example of Confucianism demonstrates, there is often a danger of oversimplification, especially since the "grand cultural theorists" often disregard endogenous processes of change occurring within each cultural tradition in response to changing incentives and threats. To take but one example, a supposedly Confucian Chinese culture has at various times and places combined with both decades of rapid economic development and centuries of stagnation within economic frameworks running the gamut from a free-market economy to centrally planned isolationism.

Our analysis in this book views economies as consisting of both slowly and rapidly changing variables with either individual or public effects, and considers an economic actor's location in time *and* in space as key factors we should always keep in mind when discussing economic processes or outcomes. We think that one of the greatest pitfalls in the history of ambitious theorizing has been the elevation of one factor at the expense of others. Wages can be flexible or rigid, but other things matter, too. Institutions are important, but they will not help much if someone is too far from the action. There may be cultural challenges, but cultures sometimes change and sometimes people even adopt a different culture.

Consider the following examples as alluding to the multi-causal reasoning that is often necessary. Britain has more flexible labor markets than France, but France has more flexible land markets. New Zealand has relatively well-functioning institutions with low levels of corruption, but it is in a rather remote location. Since 2010 Beijing has hosted the production of more peer-reviewed articles in science and engineering than any other city, but the mean number of citations per paper is much smaller than in London. The list goes on, and we think it is important to remember that we live in a multifaceted world with a complex structure of different types of capital. Capital includes not only material capital goods but also human and social capital, all of which give rise to different spatial accessibility effects. These accessibility effects in turn affect local, regional and national development trajectories.

SPACE AS AN EXPLICIT DIMENSION IN THEORIZING AND MODELING

Space was mostly absent from the work of the main economic theorists before 1940, with the notable exception of a handful of European economists such as von Thünen (1826/1930), Launhardt (1885), Weber (1909/1922), Palander (1935) and Lösch (1955). Most of these early contributors used German as their preferred language; those economists who did not read German tended to ignore them. Their theories conceived of space as the size of the land area and they modeled interactions as continuous flows in one- or two-dimensional space as seen on a map.

Most economists regarded land use, location choice, and the roles of transport and logistics as peripheral problems. The only exception was international trade theory, but that theory treated space in a rather peculiar way: trade theorists focused on the political entities of nation states. The trade of the United States (with a population of more than 320 million in a land area of 9,183,517 square kilometers in 2016) and that of a small country such as Luxembourg (563,000 residents in 2,586 square kilometers) were treated as if these size differences would be of no consequence from a theoretical or empirical point of view.

The classical economist David Ricardo was the first economist to theorize about economic phenomena with such a rudimentary conception of space. Ricardo's theory of comparative advantage is a prominent example of the discrete division of space into countries. His approach foreshadowed the development of the space-less theory of trade—with two production factors, two goods and two countries—that features in the textbook treatment of international trade to this day. Meanwhile temporal considerations have remained implicit in most of this theory, emerging only as comparative statics; there is no theory of the process that transforms initial conditions into final solutions.

In the first half of the twentieth century, Heckscher and Ohlin (Ohlin, 1933/1967) reformulated international trade and location theory so that it would also encompass the comparative supply of nationally trapped resources. Ohlin had realized that he could generalize the theory of international trade, thereby extending it to interregional trade and location choice among nations and regions. These theories and models use a discrete representation of space, which means that one can denote each of N exogenously given spatial units in the following manner: x_A for the production volume in region A, x_B for the production volume in region B and x_{AB} for the trade or other interaction between region A and region B. The main result is that equilibrium trade flows between two regions imply uniform relative prices of all goods in both regions (assuming zero transaction and transport costs). Subsequent contributions often referred to this result as the law of one price.

In the following example, we assume two regions (A and B) with identical preference structures but with different initial factor endowments and thus with different production possibility sets. The two regions are initially isolated from each other and have to rely on their own possibility sets to produce goods for their own consumption. We can then compare the consumption possibilities without and with trade between the two regions (see Figure 1.3):

1. Autarky equilibrium A^B *and* A^A (no trade implies different relative prices and that production equals consumption).
2. Trade equilibrium with $C^B = C^A$ (both countries consume the same amounts at the same relative prices). Total consumption and welfare exceeds the levels that are attainable at the production possibility frontiers without trade.

In Chapter 3, we show that the Heckscher-Ohlin and Ricardo models of trade are special cases of theories by Johann Heinrich von Thünen and Martin Beckmann. The latter theories explicitly include space and

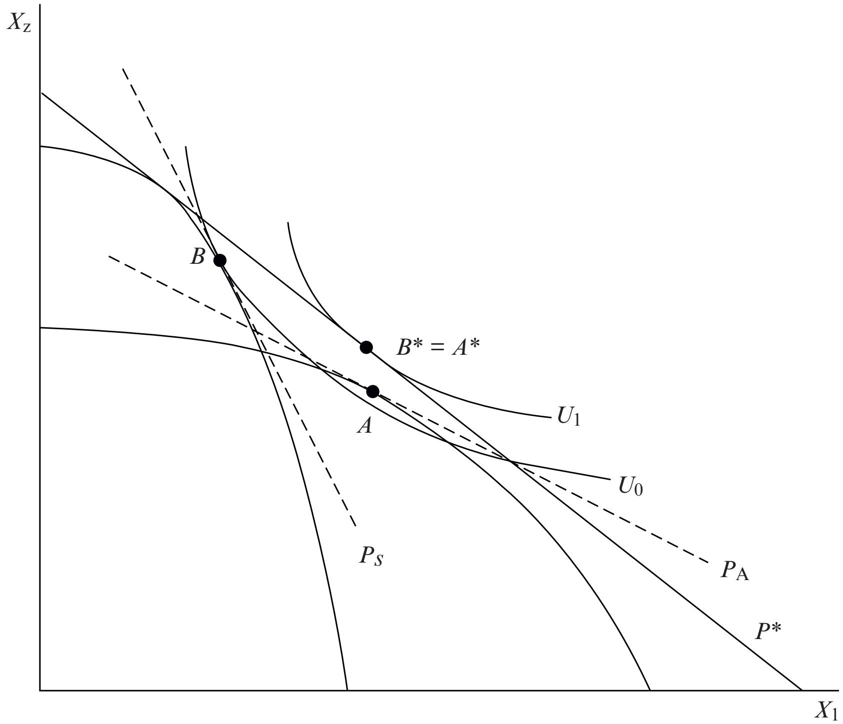

Figure 1.3 The advantages of trade for two regions (A and B) with different trapped resources

transport costs as well as production possibilities. One consequence of this more general approach is that it shows that trade is advantageous even in the case where production possibilities and consumer preferences are identical in both regions. This is a good illustration of how the explicit consideration of realistic spatial representations improves the explanatory power of a theory, and yet another instance of how theories improve when they incorporate more key variables. It is also an illustration of how there are three big blind spots in most economic theory: spatial accessibility, institutions and time-dependent processes of change. Adding one or more of these factors almost always results in an improved theory.