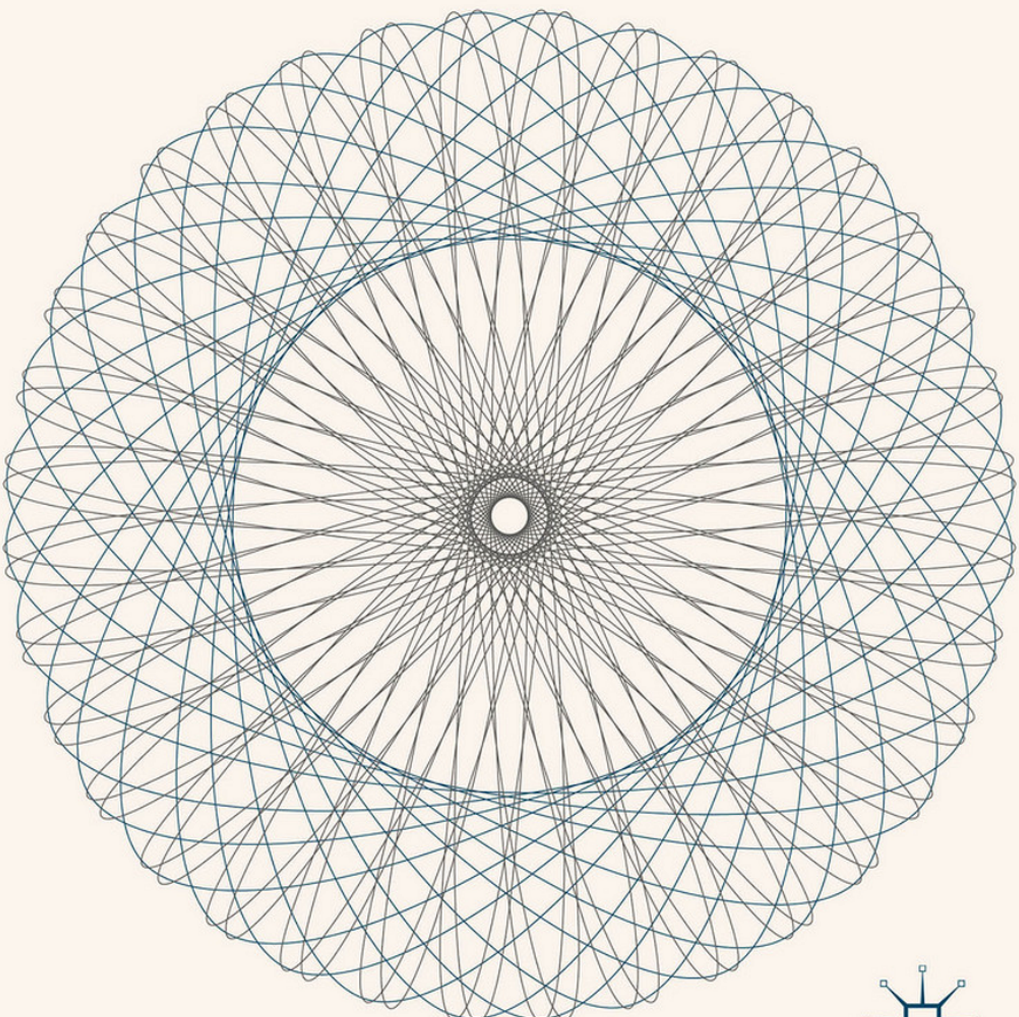
Adolfo Figueroa

Growth, Employment, Inequality, and the Environment

Unity of Knowledge in Economics
Volume I



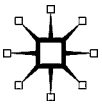
Growth, Employment, Inequality, and the Environment

Growth, Employment, Inequality, and the Environment

Unity of Knowledge in Economics: Volume I

Adolfo Figueroa

palgrave
macmillan



GROWTH, EMPLOYMENT, INEQUALITY, AND THE ENVIRONMENT

First published in 2015 by
PALGRAVE MACMILLAN®
in the United States—a division of St. Martin's Press LLC,
175 Fifth Avenue, New York, NY 10010.

Where this book is distributed in the UK, Europe and the rest of the world, this is by Palgrave Macmillan, a division of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries.

ISBN: 978–1–137–50266–7

Library of Congress Cataloging-in-Publication Data

Figueroa, Adolfo.
Growth, employment, inequality, and the environment : unity of knowledge in economics / Adolfo Figueroa.
volumes cm
Includes bibliographical references and index.
ISBN 978–1–137–50266–7 (vol. 1 : hardback : alk. paper)—
ISBN 978–1–137–50696–2 (vol. 2 : hardback : alk. paper)
1. Economics. 2. Capitalism. I. Title.

HB71.F497 2014
330—dc23 2014041730

A catalogue record of the book is available from the British Library.

Design by Newgen Knowledge Works (P) Ltd., Chennai, India.

First edition: April 2015

10 9 8 7 6 5 4 3 2 1

To Yolita Vásquez, beloved wife and great partner

Contents

Figures

Tables

Preface

Unity of knowledge is one of the fundamental epistemological requirements of science; hence, fragmentary knowledge does not imply scientific knowledge. The example of physics is very illustrative. As we know, quantum physics and relativity physics are inconsistent with each other, that is, they both cannot be true. The world of quantum physics operates with disorder, whereas the world of relativity with order. How could order in the large bodies be the result of disorder in the subatomic world? The two good partial theories do not lead to a good unified theory. The most important challenge of physics today is to find a unified theory of physics.

Economics confronts similar problems of fragmented knowledge. Good partial economic theories are inconsistent with each other and thus lack unity of knowledge. In the areas of microeconomics, macroeconomics, growth economics, and environmental economics consider the following examples. Savings is detrimental in macroeconomic theory, as more savings will reduce aggregate demand, and thus reduce total output and employment; however, savings is a social virtue in growth theory, as more savings implies higher levels of investment and output. In microeconomic theory, individuals are able to buy and sell at the market price the quantities they are willing to exchange, which implicitly assumes that individuals operate in Walrasian markets; however, macroeconomic theory assumes that some markets are non-Walrasian, such as the labor market. Finally, according to neoclassical growth theory, output can grow forever; but according to bioeconomics, there cannot be such a thing as sustainable output growth.

After two centuries of capitalist development, the world we see today is one with significant progress in the production of mountains of goods and the availability of new technologies; however, there has not been an equal social progress. The capitalist system continues to grow under social

stress, emerging from problems of employment and inequality, and lately under environmental stress. These are the fundamental social problems of our time.

Economics certainly faces the challenge of finding a unified theory to explain these features of capitalism. Standard economics has not been able to solve this challenge. Thus, there is a need for new contributions.

Main Features of the Book

The book intends to contribute to the explanation of contemporaneous capitalism under the framework of a unified theory, which seeks to provide unity of knowledge in the following areas:

(a) Production and distribution in the First World (rich countries) and in the Third World (poor countries), which constitute the basic components of the capitalist system.
(b) Economic growth and environment degradation.
(c) The economic process in the short run, long run, and the very long run.

Regarding area (a), the unified theory presented in the book will assume that the capitalist system is composed of two types of societies that differ by their income levels: the rich First World and the poor Third World. A partial theory is developed to explain production and distribution in each type of society, taken separately, followed by a unified theory that is able to explain the capitalist system, taken as a whole.

As to area (b), the economic growth process is presented, first, as a mechanical process just to understand its determinants and then as an evolutionary process. In the first case, the role of the biophysical environment is ignored, whereas it is essential in the second.

Finally, regarding area (c), the typical logical inconsistencies between microeconomics and macroeconomics will be resolved in the book by the use of general equilibrium models (non-Walrasian), in which the macrotheory of production and distribution has microfoundations; moreover, the microtheory of individual behavior has in turn macrofoundations. Therefore, the fragmented knowledge would only arise from the partial theories dealing with the different "run" categories of analysis listed above. This book will indeed construct models of different "runs" that do not contradict each other. Therefore, environmental degradation and quality of society will be explained by the very long run analysis, which will be consistent with the functioning of capitalism in the short run and long run, that is, unity of knowledge will be ensured.

In order to explain capitalism, partial theories, logically consistent with each other and constituting a unified theory, will be constructed. However, according to Popperian epistemology, theories—logically correct propositions—may turn out to be empirically false. In order to avoid this error, additional rules of scientific knowledge are needed. The book presents a method that contains such rules. The method is derived from Popperian epistemology and makes this epistemology operational. It is called the *alpha-beta method*, in which alpha constitutes the set of assumptions of the theory and beta the set of empirical predictions that is logically derived from alpha. Beta propositions make the theory empirically falsifiable or refutable.

The alpha-beta method is used throughout the book. The result is that the proposed unified theory is not refuted by the basic facts we observe in the real world. The proposed unified theory is thus empirically valid; it is a good approximation of the real world; thus, on epistemological grounds, we can accept the theory and say that scientific knowledge has been attained.

Theoretical economics is thus treated under explicit epistemological rules, in which theory is essential to attain scientific knowledge, that is, no theory no scientific knowledge; however, theory is the servant, not the master.

The Content of the Book

Part I (volume 1) lays the preliminaries of the book. The alpha-beta method together with the empirical regularities of the capitalist system are presented here to set the rules of scientific knowledge and the facts against which theories will be put up for refutation throughout the book. Then, basic models of standard economics are presented, and it is shown that some of the set of empirical regularities refute these theories. The need now arises for new partial theories and a unified theory.

The unified theory is presented in four parts. Part II (volume 1) deals with the short-run economic process. Partial economic theories of production and distribution in the First World and the Third World are developed. These partial theories seek to explain the determinants of output, employment, and distribution in each component of the capitalist system. Part I (volume 2) develops the long-run economic process, which is logically consistent with the short-run analysis. It presents the partial theories of growth and distribution in the First World and the Third World, taken separately, and then the unified theory, which explains the capitalist system taken as a whole.

Part II (volume 2) extends the analysis of the capitalist system to the very long run, in which the biophysical environment is incorporated into the economic process. It shows that output growth has three side effects.

First, growth implies a higher degree of inequality in the current generation, which in turn leads to higher degrees of social disorder. Second, growth implies the degradation of the biophysical environment, which is continuous and irrevocable, due to the physical laws of thermodynamics; hence, there are limits to economic growth: output growth cannot go on forever. Third, growth implies a higher degree of intergenerational inequality.

These side effects imply that as economic growth proceeds, the quality of society for this and future generations will eventually tend to decline. Human society is subject to increasing degrees of risk—either measurable or nonmeasurable, also called uncertainty—due to increasing degrees of social disorder and shocks from the environment on human health. Moreover, human species survival, as we know it, will tend to come to an end in a shorter span of time. These outcomes of the economic process constitute the new fundamental problems of our time. In sum, the book shows that in the very long run the economic process is evolutionary, as qualitative changes accompany the quantitative changes of the economic growth process.

Unity of knowledge implies that in this book logical inconsistencies are eliminated within each part and also between parts. We are able to construct a unified economic theory of capitalism. Because the capitalist countries have been dominant in the world output and world population, the unified theory of capitalism explains environmental degradation and the changes in the quality of human society. The available empirical regularities of capitalism do not refute the predictions of the unified theory.

Finally, Part III (volume 2) deals with the public policy implications of the unified theory. Because the unified theory is not refuted by the available empirical facts, it can be accepted as a good approximation of the real world at this stage of our research. Science-based policies can then be derived from the theory. According to the current paradigm, economic growth implies social progress; if not, what is needed for is faster economic growth. According to this view, economic growth is panacea. In contrast, according to the unified theory, growth implies social progress only in the first stages of capitalist development; later on, the side effects become significant and thus growth is subject to diminishing returns; eventually, the side effects dominate, and the net effect of growth upon social progress will be negative. Even if growth were panacea, it is ecologically unsustainable. The tradeoff between economic growth and quality of society is thus the fundamental social choice problem of our time.

In order to accomplish this formidable task in a single book, the unified theory is presented through models of a very high level of abstraction (few markets and few social actors). The study of more specific problems of capitalism will require models that are more complex.

Although the unified theory rests on new foundations, the book does not start from the scratch. From standard economics, for example, the unified theory incorporates the recent literature on growth theory and the efficiency wage theories. From outside standard economics, the book uses the work of Nicholas Georgescu-Roegen, who introduced the laws of thermodynamics into the economic process. The book will thus rely upon some relevant contributions that exist in the literature, while presenting at the same time original partial theories and the unified theory, that is, the proposed unified theory is a new scientific endeavor, and it relies on new partial theories as well. The contribution of the book is not much on new empirical findings. Its intended contribution is to provide a scientific explanation to the observed empirical facts.

In sum, this is a book about science, about the science of economics. To be sure, the book is not about further applications of the current standard economics; it is not a textbook; it is not about a survey of the literature in economics; it is not about defending or attacking certain economic doctrines either. The book seeks to push forward the discipline of economics into new areas of enquiry—those that constitute the fundamental social problems of our time—by using the rules of scientific knowledge derived from Popperian epistemology. The book intends to contribute to the progress of economic science in both scope and method.

Intended Readership

The readers of the book are required to have a basic knowledge of standard economics. This is just in accord with the principle that scientific progress is a cumulative process. Basic knowledge of mathematics is also needed. However, the reader is expected to think economics—not mathematics.

The book is written primarily for advanced undergraduates and first-year graduate students in economics. It can be used as complementary reading in courses of standard economic theories (macro, micro, labor, growth, distribution, and environment economics) to expose students to different views on method and scope in economics and to different partial theories as well. Those interested in conducting seminars and research seminars would also find the book very useful due to the new areas of inquiry and the new research methodology (alpha-beta method) that it presents. The development of partial theories about the First World and the Third World, together with the unified theory, should attract the attention of students of both regions, that is, students should find the book meaningful for learning more not only about their own reality, but also about the capitalist system taken as a whole.

Scholars in disciplines other than economics should also find the book useful. Political scientists may be interested in the role of democracy in the functioning of capitalism; sociologists in the relation between inequality, social disorder, and quality of society; and physicists, biologists, and ecologists in the interrelationships between the economic growth process and the environment, presented analytically as the intergenerational consumption frontier. For those working in basic research in the social sciences and natural sciences, the findings of the book may suggest research topics in particular settings.

Professionals working in applied research should also find the book helpful because new testable hypotheses on specific problems or places can be derived from it. The book concludes with a discussion of science-based public policies. These results should be of interest of those doing policy-oriented research.

The main findings of the book should also be accessible to general readers interested in knowing what is new in the science of economics that is relevant for the fundamental social problems of our time. The reason is that theories are presented through very simple models in which very simple abstract societies are constructed, in which only the essential factors of the complex real world are taken into account; the empirical refutation of theories is presented by just summarizing the relevant statistical and econometric works; the graphs are intended to illustrate the basic relations and make the theories more accessible; finally, the conclusions and the way they have been reached are stated in simple language at the end of each chapter. Thus, general readers will know what is going on at each stage of the construction of the unified theory; moreover, the public policy implications of the unified theory (chapter 10, volume 2) and its contrast with the current paradigm should then be accessible to the general public.

Acknowledgments

Some words of gratitude are in order. The main idea about the unified theory started around 1997 at the University of Texas-Austin, and then was continued at the Catholic University of Peru, in the Economics Department, until 2007. Thereafter, research has been carried out at Centrum Graduate Business School of this university, where the support of its director Fernando D'Alessio and the research assistance of Erick Vila, Francisco Pardo, Javier Vásquez, and Cynthia Paz contributed greatly to the completion of the book. Participating in collaborative research projects with the University of Wisconsin-Madison in the United States, and the University of Bath and University of Oxford in the United Kingdom, and interacting with Professors Michael Carter, James Copestake, Frances Stewart, and Rosemary Thorp, was particularly important in the development of the theory. Parts of this book were used as teaching material in courses and seminars in economics at the University of Texas-Austin, University of Wisconsin-Madison, and at the Catholic University of Peru, where professors and students contributed to the development of the unified theory.

Adolfo Figueroa

PART I

Scientific Knowledge Rules, Facts, and Standard Theories Refutation

CHAPTER 1

Popperian Epistemology in Economics: The Alpha-Beta Method

Why has the progress of scientific knowledge in economics proceeded at a pace that is slower than that of the natural sciences? A possible reason is the limitation of data, in quantity and quality; in addition, the instruments of measurement for the social phenomena are relatively imperfect. The other reason seems to rest upon the role of methodology in the construction of scientific knowledge in economics. Compared to physics, economics seeks to explain the functioning of the social world, which is a much more complex world than the physical world; one may then propose the principle that understanding more complex worlds, such as the social world, is more demanding on methodology than understanding the physical world.

Economist Paul Samuelson wrote in his classic book *Foundations of Economic Analysis* about this principle as follows:

> [This] book may hold some interest for the reader who is curious about the methodology of the social sciences . . . [I]n a hard, exact science [as physics] a practitioner does not really have to know much about methodology. Indeed, even if he is a definitely misguided methodologist, the subject itself has a self-cleansing property which renders harmless his aberrations. By contrast, a scholar in economics who is fundamentally confused concerning [methodology] may spend a lifetime shadow-boxing with reality. In a sense, therefore, in order to earn his daily bread as a fruitful contributor to knowledge, the practitioner of an intermediately hard science like economics must come to terms with methodological problems. (1947, pp. viii–ix)

If the separation between physics and economics made by Samuelson as hard and intermediately hard sciences was transformed into complex (economics) and less complex (physics), as will be done in this book, then the principle proposed above would follow. Economics is a more methodology-intensive science than physics.

Methodology is another name for epistemology (from the Greek *episteme*, knowledge). Epistemology deals with the logic of scientific knowledge, from which a practical set of rules to arrive at scientific knowledge can be derived. Such a set of rules is needed in economics. Popperian epistemology will be used for this purpose, from which the set of scientific rules—called the *alpha-beta method*—will be derived. The method is presented in this chapter and will then be applied in the entire book. (This chapter draws heavily on Figueroa [2012].)

Deriving Scientific Rules from Popperian Epistemology

Scientific knowledge seeks to establish relations among objects and explain them. The objects can be mental or physical. Formal sciences study the relations among mental objects, whereas factual sciences study the relations among material objects. Mathematics and logic are examples of formal sciences; physics and economics are instances of factual sciences.

Scientific knowledge can be seen as a set of propositions that is error free. What would be the criterion to accept or reject a proposition as scientific? It depends upon the type of science. In the formal sciences, the criterion seems to be rather straightforward: the relations established must be free of internal logical contradictions. In the factual sciences, by contrast, the criteria are more involved. The propositions of a factual science must be free of internal logical contradictions as well. However, this criterion constitutes just a necessary condition; empirical consistency between the propositions and the facts will also be required. The real world cannot be explained by using deductive logic alone. This is the logic that corresponds to formal sciences, not to factual sciences.

According to the epistemology developed by Karl Popper (1968), scientific knowledge that seeks to explain the real world cannot be attained by using inductive logic. There is no such thing as inductive logic, that is, there is no logical way to go from particular empirical observations to general relations. His classical example is: no matter how many instances of white swans we may have observed, this does not justify the conclusion that all swans are white.

The logic of scientific knowledge developed by Popper can be summarized as follows: Theory (an abstract world) is needed to explain the real

world; from theory, some conclusions about reality are derived by logical deduction. These conclusions, the empirical predictions of the theory, can then be submitted for confrontation against reality. If the empirical predictions and reality are consistent, we have no reason to discard the theory; if they are inconsistent, then the theory has been proven false and it has been falsified. Hence, a scientific theory must be able to generate falsifiable or refutable empirical propositions.

An empirical proposition is falsifiable if in principle it can be false. This is the *principle of falsification*. For example, the proposition "It will rain or not rain here tomorrow" is not falsifiable, whereas the proposition "It will rain here tomorrow" is. Falsification is therefore the demarcation principle between scientific and nonscientific propositions.

The basic scientific rules that can be logically derived from Popperian epistemology include:

(a) Scientific theory is required to explain the real world. No theory, no explanation.
(b) Falsification is the criterion of demarcation. The scientific theory must be falsifiable in the sense that the empirical propositions derived from it, by deductive logic, must be, in principle, false.
(c) The theory is rejected if its empirical predictions are refuted by facts; if they are not, the theory is accepted provisionally until new data or superior scientific theory appears.

The question now is to see whether these scientific rules can be applied to economics. In order to answer this question, we must first be clear about the scope of economics.

The Economic Process

Economics is a social science. It seeks to explain a particular aspect of the functioning of human societies: the determinants of the production of goods and its distribution between social groups. Economics studies those social relations that are related to the production and distribution of goods. Goods constitute the cement that links people in social relations. This is the standard scope of economics.

Human societies constitute complex realities. The notion of complexity can be defined by the existence of a large number of elements and the heterogeneity among them that forms the reality, together with the multiple factors that shape the relations between those elements. Human diversity, together with the multiplicity of human interactions, makes human

societies intricate realities. The simple fact that individuals in a human society are not identical, as compared to the homogeneity of atoms in the physical world, suggests that the social world is more complex than the physical world. Human societies are complex systems of interacting individuals in which individuals themselves are complex systems.

How can a complex social reality be subject to scientific knowledge? The usual answer is to use the method of abstraction. This method rests on two particular assumptions. The first assumption is that the complex social reality can be transformed into a process, in which the social relations occur regularly and repeatedly. The second is that the complex social reality is reducible to a simpler abstract world by setting aside elements of the process that are not important to understand the social world.

The concept of process to be used in this book refers to the one developed by economist Nicholas Georgescu-Roegen (1971, Chapter 9). A process is as a series of social activities carried out in parts of the real world, having a given duration (a given unit of time) and a purpose, and repeated period after period. The essential characteristics of a process then include the existence of a boundary that separates the outside world from the inside because the process refers to a partial aspect of social reality. Moreover, there are elements that cross the boundary from outside the process—called the *exogenous* elements—and those that cross the boundary from inside the process—the *endogenous* elements. In a process, there is also an underlying mechanism by which the exogenous elements influence the outcome of the endogenous elements. The other characteristic is repetition: process always implies repetition in which the unit of time is well defined.

The analytical transformation of a complex social reality into a process implies the separation of all elements of reality into endogenous and exogenous. The complete list of endogenous and exogenous elements of a process would include observable and nonobservable elements. When observable, call these elements *endogenous variables* and *exogenous variables*. The method of abstraction must now be applied to reduce the full process to a simpler process, which is called *process analysis*. The use of abstraction implies the selection of the most significant endogenous and exogenous variables together with the most significant mechanisms of the process.

Certainly, to present the full process, with a complete list of the variables, would be equivalent to constructing a map of reality to the scale 1:1. As in the case of the map, a complex reality cannot be understood at this scale of representation. Abstraction implies that some variables and some mechanisms must be ignored. This is how a real world is transformed into an abstract process, into an abstract world, in which only the supposedly important variables and mechanisms are included and the rest are just ignored.

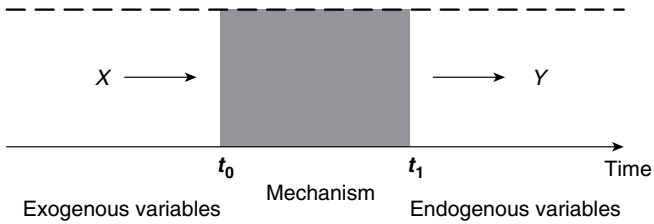


Figure 1.1 Diagrammatic representation of process analysis.

Figure 1.1 shows the analytical representation of process analysis. The segment t_o–t_1 represents the duration of the process (the unit of time), which is going to be repeated period after period. *X* is the set of exogenous variables and *Y* is the set of endogenous variables. The shaded area indicates the underlying mechanism by which *X* and *Y* are connected.

What happens inside the process is unobservable, as indicated by the shaded area in the figure. If it were observable, the interior of the process would have to be considered as another process in itself, with other endogenous and exogenous variables and another mechanism; but the latter mechanism would also be observable and then constitute another process, and so on. Thus, we would arrive at the logical problem of an infinite regress. (Science avoids this trap by making assumptions about the initial conditions of a process.) Ultimately, there must be something hidden beneath the things we observe, which are contained in the mechanisms of the process. Science seeks to unravel those underlying factors.

Because a process repeats itself period after period, the relationships between exogenous and endogenous variables can be observed continuously and thereby systematic relationships or empirical regularities can be observed. The existence of empirical regularities is a necessary condition for scientific knowledge. A chaotic world—where regularities are absent—is much harder to understand; therefore, it will not be a part of process analysis.

How do we decide which elements are important in a process and which are not? This is established by a set of assumptions, that is, by constructing a *scientific theory*. A scientific theory is a set of assumptions about the workings of the abstract world, which intends to resemble well the real world.

The endogenous variables constitute the object of study, the phenomena that the theory intends to explain. The exogenous variables are the explanatory factors. The construction of an abstract process is made through the introduction of assumptions about which exogenous variables of the process are important and which are not. In addition, assumptions must be made about what the relevant underlying mechanisms of the process are.

Hence, the theory determines the endogenous and exogenous variables and the particular mechanisms of the abstract world. A theory is, therefore, a logical artifice by which a complex real world is transformed into a simple abstract world. This abstract world is much simpler to understand than the real world.

A method to transform the real social world, a complex real world, into an abstract world is therefore needed. This method must be consistent with both Popperian epistemology and with process analysis. The alpha-beta method constitutes such a method, which is now presented.

The Alpha-Beta Method

The alpha-beta method starts with the following definition:

> In terms of the logical ordering of its propositions, a scientific method can be seen as a set of alpha and beta propositions, in which beta propositions are logically derived from the alpha and no alpha proposition is derived from another alpha; hence, alpha propositions constitute the primary assumptions of the theory. (adapted from Georgescu-Roegen 1971, p. 26)

A theory is a set of assumptions, as mentioned earlier. Include now the condition that the elements of the set must be logically consistent. Then *scientific theory is a set of logically consistent assumptions about the functioning of the real world*. Alpha is that set of assumptions of the theory. As indicated above, the assumptions refer to the components of a process: the endogenous variables, the supposedly significant exogenous variables, and the supposedly significant underlying mechanisms that connect them. With these assumptions, an abstract world can be constructed.

Can the assumptions of a theory be derived from empirical observations? No, they cannot. The reason is that the theory seeks to explain those empirical observations. What we can get from reality by empirical observation is a description of it, not an abstraction. The listing of all elements of the process by itself cannot lead to the discovery of the essential and nonessential elements. Thus, the assumptions cannot intend to reproduce reality, but must ignore some elements of reality. This is the use of abstraction.

The assumptions of a theory are in the nature of axioms, in the sense that they need no justification. If the set of assumptions needed justification, another set of assumptions to justify them would be needed, which in turn would need another set to justify the latter, and so on; hence, we would

reach the logical problem of infinite regress. Therefore, the assumptions of a theory cannot have justification. They are established arbitrarily; they are gratis, and thus constitute a theoretical hypothesis, which may be empirically true or false.

However, the alpha propositions must satisfy the following logical requirements to be part of the Popperian epistemology. First, they must be unobservable, as they refer to the underlying mechanisms connecting the endogenous and exogenous variables, such as assumptions about the motivations that guide the actions of social actors. Second, they must also be nontautological, that is, in principle, the assumptions could be false. For example, the assumption that everyone acts as he or she desires is unobservable but tautological; any observed behavior would be consistent with it; however, the assumption that firms act guided by the motivation of profit maximization is unobservable and nontautological, for it may be empirically false. Third, alpha propositions must constitute a logical system in which assumptions should not contradict each other.

Why is falsification of a theory needed? The reason is that the set of assumptions of the theory was established arbitrarily, as a theoretical hypothesis—only subject to the conformation of being a logical system. Thus, alpha propositions correspond to the initial step in the construction of a good or valid theory, that is, this set of assumptions is part of an algorithm. According to Popperian epistemology, a theory is constructed as part of an algorithm, of a trial-and-error process, the aim of which is to reach a good theory. A good theory is one that has constructed a simple abstract world that resembles well the real world. Thus if the initial theory fails empirically, a new set of assumptions is established to form a new theory, and a new abstract world is thus constructed; if this second abstract world does not resemble the real world, the theory fails and is abandoned, and a new set of assumptions is established, and so on. The good theory is found by a procedure of trial and error in which we assist in the funeral of wrong theories. Popperian epistemology is evolutionary.

An implication of the falsification principle is that theories must be mortal, that is, they must be falsifiable. Hence, what the assumptions of a theory need is not justification, but falsification against data of the real world, independent of how, why, or by whom the assumptions were chosen. In this sense, Popperian epistemology leads to critical scientific knowledge.

How is the falsification of a theory made? Not through alpha propositions directly, as they are unobservable, but indirectly through beta propositions. It is here that beta propositions play a significant role. Beta propositions are derived logically from the set of alpha propositions. The logical derivation is about the relations between the endogenous and exogenous variables

of the theory. Therefore, three characteristics of beta propositions can be established:

- Beta propositions constitute the empirical predictions of the theory, that is, the alpha-beta method transforms nonobservable propositions into observable ones.
- Beta propositions are *falsifiable*. The relations between exogenous and endogenous variables that beta propositions derive from the theory (a set of nontautological propositions) are observable and refutable. Hence, beta propositions make the theory falsifiable: if a beta proposition fails, then the theory necessarily fails.
- Beta propositions show the *causality* relations, that is, the relations between the exogenous variables and the endogenous variables of the theory. Changes in the exogenous variables are the cause and changes in the endogenous variables are the effect. Note that causality requires a theory—no theory, no causality.

On the last characteristic of falsifiability, note that the beta propositions may or may not coincide with what we observe in the real world. *Although a beta proposition is logically derived from a theory, which in turn is a logically correct system, it may be empirically false*. The reason is that the set of assumptions contained in the alpha propositions were selected arbitrarily, for there is no other way to select them.

From the way beta propositions are derived, it follows that for each endogenous variable of the theory there will be a beta proposition; hence, there will be as many beta propositions as there are endogenous variables in the theory. Alternatively, we can say that beta propositions show the causality relations for each endogenous variable.

A scientific theory must therefore be falsifiable through its beta propositions, that is, in principle, a scientific theory must be mortal. "Men die when God so wishes" is an example of a nonfalsifiable proposition. The reason is simple: God's wishes are unobservable. Shamans are particularly good at using nonfalsifiable propositions to deal with sick people: "If you have faith on my medicine you will get well." If the person complains that he is not getting well, the shaman could say, "You had no faith in my medicine." This statement never fails because faith is unobservable. Apart from unobservable elements, there is another case in which a proposition is nonfalsifiable: when it predicts all possible outcomes ("It will rain or not rain here tomorrow"). These types of propositions are tautological, useless to falsify a theory. These examples illustrate the principle that a "theory" that cannot generate beta propositions becomes immortal and cannot be a scientific theory.

How do we decide to accept or reject a theory? If all beta propositions coincide with reality, the theory is not refuted by the available facts; if at least one beta proposition fails, the theory fails to explain the reality. Beta propositions represent the necessary conditions to accept the theory as valid, that is, if a single beta proposition fails, the theory necessarily fails because the beta proposition was logically derived from the set of assumptions of the theory.

The last principle can be illustrated with a simple example. Consider a theory that states "Figure F is a square" (suppose the Figure F is unobservable for the researcher). By definition, a square is a rectangle with all four sides equal. If this is taken as the assumption of the theory, then the following beta proposition can be logically derived: the two diagonals must be equal. If the available empirical evidence shows that the diagonals are not equal, Figure F cannot be a square. The theory is refuted by this fact. However, if empirically the diagonals are equal, we can only say that the theory is *consistent* with facts: We cannot conclude that F is a square, for it could also be a rectangle (with adjacent sides of unequal length). As this example shows, a theory can only be *corroborated* by facts; it cannot be *verified*, proven to be true. Truth is much harder to establish in science.

Which particular assumptions are responsible for the failure of a theory? It is not possible to identify any particular assumption that is responsible for the failure of a theory. We know that a beta proposition is logically derived from the entire set of assumptions of a theory. Due to this *problem of identification*, a good theory cannot be derived logically from the old one, but it would have to be invented, by trial and error, using new sets of assumptions.

To understand a complex world the use of theory is a necessity. As shown earlier, theory is an abstraction of the real world, in which only some elements of the process are taken into account. The beta propositions constitute the empirical predictions of the theory and are utilized to seek refutation of the theory. Logically, therefore, a beta proposition can fit only the general or typical cases of real-world observations. Due to the use of abstraction, it may not fit all observed cases. Therefore, the refutation of a theory needs to be based on statistical principles.

Due to the use of abstraction, the empirical prediction of a theory must then be confronted against the relationship between the average values of the observed variables. If a theory on income inequality predicts that incomes of individuals depend directly upon their years of schooling, this prediction cannot be refuted just because an individual with no education has a high income, for he or she may be the exception. The theory will be refuted if the *average* incomes of people are inversely related to *average*

years of schooling, which refutes the general relation established by the theory.

A single empirical observation that contradicts a beta proposition is insufficient to refute a theory, for the statistical value of one observation is nil. This observation could just correspond to an exception, a deviation from the average due to other factors that the theory ignored (use of abstraction), or purely by chance, which is called *statistical error*. A single counterexample is sufficient to invalidate a theorem in mathematics, but it is not sufficient to refute a scientific theory. Accordingly, a distinction must be made between *statistical error* of a theory and *failure* of a theory.

The assertion that a theory explains reality therefore has a precise meaning: its beta propositions are statistically *consistent* with empirical data. On the other hand, if empirical data do not fit the beta propositions, the theory is simply *false*. In this case, a new theory should be formulated and the algorithm continued.

The continuous interaction between theory and empirical data is the cornerstone of the alpha-beta method. The logic of this method is thus clear. A complex reality is reduced to an abstract process—where there is repetition, and where regularities in empirical relationships can occur—with the use of a theory (alpha propositions); from the theory, empirical predictions are logically derived (beta propositions), which are falsifiable; and the falsification is made through statistical analysis. If the beta propositions are statistically consistent with facts, there is no reason to reject the theory, and it is accepted as valid; if the beta propositions are statistically refuted by facts, then the theory is rejected and, thus, a new theory is constructed. Scientific knowledge is achieved through this algorithm.

Figure 1.2 presents diagrammatically the alpha-beta method. From the set of alpha propositions α_1, the set of beta propositions, β_1, is logically derived (indicated by the double arrow). The set β_1 must then be subject to statistical testing (indicated by the single arrow). While the double arrow indicates deductive logic, the single arrow indicates operational procedure, or the task to be performed. Statistical testing of the theory implies seeking a statistical consistency between beta propositions and the available set of statistical relations or associations between endogenous and exogenous variables, which is represented by the letter **b**. This confrontation is indicated by the double-swung dash symbol (≈), which in this case means seeking the statistical confrontation between β_1 and **b**. If statistically (not mathematically) we get β_1=**b**, then α_1 is consistent with reality; there is no reason to reject the theory, and then we may accept it. If statistically we get $\beta_1 \neq$**b**, then reality refutes the theory α_1, and another theory α_2 should be developed; thus the algorithm is continued.

$\alpha_1 \Rightarrow \beta_1$ $[\beta_1 \approx \mathbf{b}]$

If $\beta_1 = \mathbf{b}$, α_1 is consistent with **b** and explains reality

If $\beta_1 \neq \mathbf{b}$, $\alpha 1$ does not explain reality and is refuted by facts. Then,

$\alpha_2 \Rightarrow \beta_2$ $[\beta_2 \approx \mathbf{b}]$

If… (continue with the same algorithm)

Figure 1.2 The alpha-beta method.

In the case of falsifying several theories at the same time, given data set **b**, some theories will be false and some will be consistent. Those theories that survive the entire process of falsification will become the valid theories. They will reign until new data set, new statistical testing methods, or a superior theory appears. A theory is superior to the others if it generates the same beta propositions as the others, but in addition generates other beta propositions that are consistent with facts, which the other theories cannot. A theory is thus superior to others when it can explain the same facts that the others can and also some additional facts that the others cannot.

It is clear from the alpha-beta method that data alone, without theory, cannot explain the real world. Call H-hypothesis the one that is not derived from a theory; then [H ≈ **b**] represents the operation of statistical testing, which if accepted will show the existence of statistical correlation among the variables included in H. However, there is no logical route from correlation to causality, that is, from H to α, no matter how sophisticated the statistical technique is. *The existence of correlation does not imply causality*. The logical route can only go from alpha to beta propositions, as shown in figure 1.2. Causality requires an underlying theory because exogenous and endogenous variables are contained in beta propositions, which can only come from a theory. Reality can only be understood in the light of a theory.

The idea that theory is not needed to obtain causality is contained in the common argument "empirical data should speak for themselves," which has no epistemological justification. Statistical testing of a beta-hypothesis leads to scientific knowledge, but that of H-hypothesis does not. From H-hypothesis statistically accepted [H=**b**], we must seek to find an alpha proposition not by deductive logic, but by trial and error, and such that β=H can be derived from it.

In sum, the alpha-beta method is in accord with Popperian epistemology. The rules of the alpha-beta method are consistent with the three rules

of Popperian epistemology mentioned earlier. These consistencies can be written as follows:

(a) The rule that theory is needed for scientific explanation operates through alpha propositions.
(b) The rule that falsification is the criterion of demarcation and that theory must be falsifiable operates through beta propositions, which are logically derived from alpha; if beta fails, then alpha fails to explain the real world.
(c) The rule that rejection-acceptance of a theory comes from the process of falsification, its confrontation with facts, operates through the statistical testing of beta propositions against empirical data.

Therefore, the alpha-beta method is a particular method of Popperian methodology. Moreover, while Popperian epistemology gives us only general rules of scientific knowledge, the alpha-beta method makes it operational in complex sciences, in which the use of abstraction needs a process diagram. Thus, Popperian epistemology, and the alpha-beta method derived from it, is in principle applicable to economics. The main features of this application will now be developed.

The Alpha-Beta Method in Economics

The scope of economics can now be defined in terms of process analysis. Figure 1.1 can be used for this purpose. Economics is a social science that studies the economic process, which is defined as the process of production of goods and the distribution of those goods among social groups in human societies. Production and distribution are the endogenous variables of any economic theory. The factors that constitute the ultimate cause of changes in production and distribution—the exogenous variables—and the underlying mechanisms are particular to each theory. An economic theory contains a set of primary assumptions from which beta propositions are logically derived. There are however some particularities in the application of the alpha-beta method to economics, which are now spelled out.

Economics Assumes No Ontological Universalism

Physics is considered the exemplar of factual sciences. A characteristic of physics is that it assumes ontological universalism, that is, the physical world is seen as a single world. Thus, the atom is expected to show the same

behavior in any place on Earth. Therefore, the theories of physics are supposed to be valid for any place and time on Earth.

As a social science, economics assumes no ontological universalism: The mechanisms underlying production and distribution are not the same for all types of human societies, independent of time and space. Then economics does not pursue a single theory that is valid for all human societies, independent of their place and time of existence. The other extreme assumption, that an economic theory can only be valid for a particular society, cannot be accepted either because economics cannot explain individual societies, only groups of societies.

Consider instead two types of assumptions on the economic process: one that is universal, common to all human societies, and the other that is specific to particular types of societies. The universal assumptions are presented here, whereas the specific assumptions will refer to capitalist societies, our subject of study, and will be presented in the rest of the book. These two levels of assumptions must form a logical system.

On the universal assumptions, consider the following: the nature of the economic problem is the same in all societies in the sense that human necessities are unlimited whereas society's capacity to produce goods in any period of time is limited. Production capacity is limited by three constraints: (1) factor endowments, which include labor, capital, land, and other natural resources; (2) technological knowledge of society, which refers to the knowledge on how to produce goods; and (3) physical laws about matter and energy relationships (laws of thermodynamics) in the transformation of natural resources into goods. Given these constraints, it is assumed that the production frontier of society in any period of time is determined. This frontier is the boundary that separates the menu of goods that society can produce, with those three constraints, from those that cannot.

In order to solve the economic problem, societies seek to establish the rules under which the economic process must operate. Economics assumes that societies operate with given institutions, which include both social organizations and the social norms of the economic process (North, 1990). As in a football game, the economic process needs organizations, that is, social actors that carry out the economic activity; it also needs a set of rules under which these agents will interact. Economic theories must then make assumptions about what the essential components of institutions in society are.

Another universal assumption is that social actors have an economic rationality. The logic underlying the behavior of people is called *rationality*. This logic refers to the motivations that guide the actions of social actors, which constitutes the forces that move society. Economics assumes that individuals

are rational in the sense that they act guided by particular motivations and aims, which are shaped by the social norms of the society they live in, that is, their rationality cannot be independent of the institutional context. Finally, a universal assumption about the initial conditions of a particular type of society is needed. This assumption makes possible the use of process analysis, including the role of history in understanding the economic process of different types of societies.

The universal alpha propositions can be summarized by a set of postulates, as follows:

- α_0 *(1). The scarcity postulate*: Human societies face the economic problem: the maximum flow of goods that society can produce is *limited* by the endowments of capital, labor, and natural resources, and technological knowledge, whereas the quantity of goods desired in society is *unlimited*. Nirvana societies do not exist, or they are the exception.
- α_0 *(2). The institutional postulate*: In order to solve the economic problem, societies seek to establish a particular institution, that is, a set of rules and organizations, which regulates the social interactions in the process of production and distribution. Different types of human societies imply different institutions.
- α_0 *(3). The rationality postulate*: Individuals act guided by motivations, which are shaped by the institutional context of the society they live in.
- α_0 *(4). The initial conditions postulate*: Economic processes of human societies differ according to their origin (history), in which two initial conditions are essential: the factor endowments and the distribution of economic and social assets among individuals. The first refers to the aggregate factor endowments, which determine labor productivity levels in society; the second refers to the initial degree of inequality in society.

These four postulates constitute a logical system, in which internal contradictions between propositions are absent.

These universal assumptions are called *postulates* just to indicate that they refer to the basic primary assumptions of the economic process. Axiom would be too rigid a term to be applied to assumptions when they are expected to play the role of a logical artifice in the construction of scientific theories, which is an iterative endeavor. *Assumptions* (more flexible than postulates) will be the term used when dealing with society-specific economic theories.

An economic theory that seeks to explain production and distribution in a particular type of human society will still consist of alpha and beta

propositions. But the alpha propositions will have two components: the universal assumptions—presented above and labeled with the zero subscript—and the society-specific assumptions, which will have a subscript to identify the particular type of society under study, such as (C) for capitalist societies, as will be the case in this book. If economics pursues to be a social science that studies the process of production and distribution in human societies, the society-specific assumptions must be logically consistent with the universal assumptions presented above.

In light of the universal assumptions, different types of human societies differ in their institutions. The most common definition of capitalist societies, and the one that will be used in this book, can be presented as follows:

> *Capitalist societies* operate under the following institutional context: (a) Rules: People participating in the economic process are endowed with economic and political assets; economic assets are subject to private property rights; people exchange goods subject to the norms of market exchange, which include the labor market. The political regime is democratic. (b) Organizations: include private firms, households, and the government.

This is the conceptual category of capitalism that will be used in the book. Economic theories of capitalism differ in their assumptions about the mechanisms of the market system and the democratic system and also on their assumptions about their initial conditions.

The Existence of Equilibrium Is Fundamental for Falsification

By definition, the outcome of the economic process will be repeated period after period. However, there are different logical ways in which the economic process can be repeated. In order to determine those forms, a definition of equilibrium in the economic process is needed. This can be stated as follows:

> The outcome of an economic process is said to be in an *equilibrium* situation, if, and only if, no social actor has both the power and the incentive to modify that outcome.

In the economic process there exist three types of relations between endogenous and exogenous variables. The first refers to equilibrium conditions. The values of the endogenous variables can be restricted to take values within a given range of values under any equilibrium situation, that

is, independent of the values of the exogenous values (e.g., "The quantity demanded must be equal to the quantity supplied in the market.") The second refers to interactions between endogenous and exogenous variables within the process, through the mechanisms of the process, which are called *structural relations or structural equations*. The third refers to relations showing the effect of changes in the exogenous variables upon the endogenous variables, as the final outcome of the economic process, which is called *reduced form relations or reduced form equations*. Therefore, the reduced form equations of the economic process show the beta propositions or causality relations of the theory.

Depending on the nature of equilibrium, three types of economic processes can be distinguished. They are: static, dynamic, and evolutionary.

First, let us consider a static process. Under *static equilibrium,* the values of the endogenous variables are repeated period after period, as long as the values of the exogenous variables remain constant. Static equilibrium may be stable or unstable. It is stable when the value of the endogenous variable spontaneously restores its equilibrium position whenever it falls out of equilibrium (the classical metaphor is a ball inside a bowl); otherwise, the equilibrium is unstable (a ball on top of a bowl that is placed upside down). Stable equilibrium implies that the system is self-regulated. Changes in exogenous variables will cause changes in the equilibrium position of each endogenous variables and thus in their values of equilibrium. The direction of the effects (positive, negative, no effect, undetermined) constitutes beta propositions. Thus, beta propositions are determined by using the method of *comparative statics*. This method requires a stable equilibrium: the endogenous variable must move from the initial equilibrium to the new equilibrium spontaneously. However, the change in an endogenous variable from one position of the equilibrium to another is without regard to the transitional process (the trajectory) involved in the adjustment (Samuelson 1947, p. 8).

A dynamic process occurs when there are intertemporal relationships among endogenous variables in the structural equations of the economic process. Hence, given the initial conditions of the endogenous variables, and the values of the exogenous variables, the equilibrium values of the endogenous variables of today determine their equilibrium values of tomorrow, which in turn determine their equilibrium values of the following period, and so on. Endogenous variables change just with the passage of time. Thus, *dynamic equilibrium* implies a particular trajectory over time of each endogenous variable, as long as the values of the exogenous variables remain fixed.

Under certain conditions, the dynamic equilibrium is just the sequence of static equilibrium situations. Therefore, if the static equilibrium is stable, the dynamic equilibrium will be too. If, for some reason, an endogenous

variable is out of the equilibrium trajectory, it will move spontaneously to the equilibrium trajectory. This adjustment will imply a particular transitional trajectory as well, which is called *transition dynamics*. In the dynamic process, beta propositions are determined by using the method of *comparative dynamics*, which shows the effect of changes in the exogenous variables upon changes in the equilibrium trajectory, including the transition dynamics involved in the adjustment.

The economic process can take the form of either static or dynamic, depending on the assumptions of the theory. Hence, when applied to economics, the representation of a process, as shown in figure 1.1, can be understood either as a static process or as a dynamic process. In the static case, the value of Y will remain unchanged if the value of X remains fixed, and Y will change to another value if, and only if, the value of X changes. In the dynamic case, the values of Y will move along a given trajectory over time if the value of X remains fixed, and the values of Y will shift to another trajectory if, and only if, the value of X changes. Static or dynamic process assumes that the economic process is a mechanical one. No qualitative changes take place in this process that may disrupt its equilibrium situation. Therefore, static or dynamic processes can be repeated forever.

If the economic process is viewed differently, as subject to qualitative changes as it is repeated, then we have an *evolutionary economic process.* In this case, the assumption is that, as the economic process is repeated period after period along a trajectory, qualitative changes will also take place, which will eventually set limits to the repetition; then a threshold value of the endogenous value will exist (Y^*). Once the threshold value is reached, a new set of relations within the process will appear, a *regime switching* will take place; hence, the endogenous variable will move to another trajectory. The economic process cannot be repeated forever, but only for a finite number of periods. The evolutionary process can still be represented in a process diagram, as in figure 1.1, but only temporarily, subject to the condition $Y < Y^*$. Evolutionary processes are more complex than mechanical ones.

Figure 1.3 presents the three types of economic processes. The reading is as follows:

1. Panel (a) shows the static process. Equilibrium of endogenous variable Y is repeated over time along the horizontal line AA′, for a given value of the exogenous variable X. At time t', the exogenous variable increases, and the new equilibrium is given by the line BB′; then at period t', point A′ is out of equilibrium and, because the equilibrium is stable, it will shift to point B spontaneously and immediately.

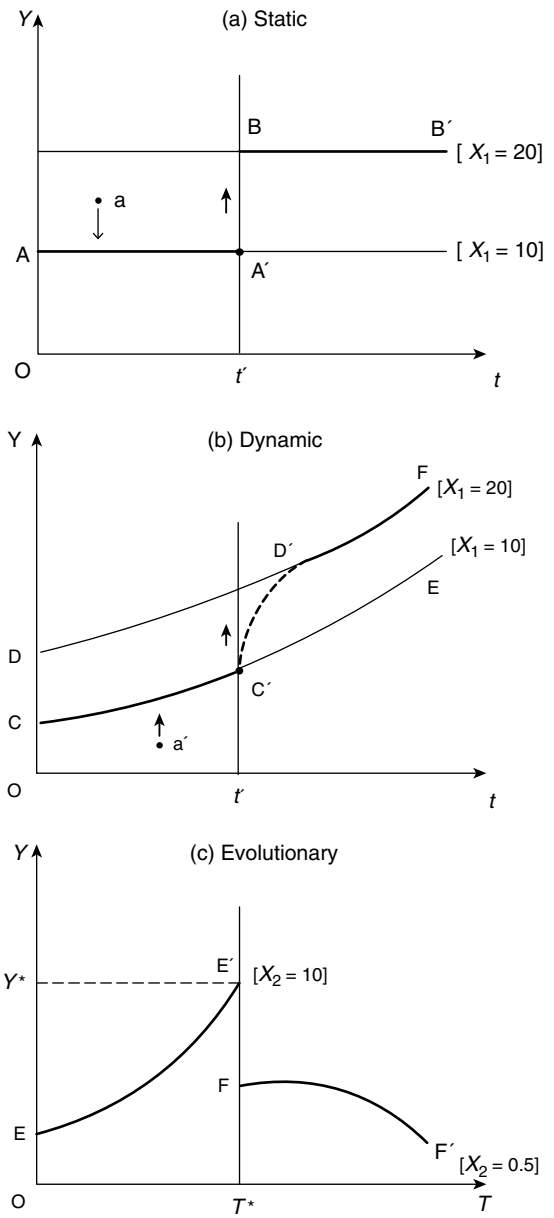


Figure 1.3 Types of economic processes: Static, dynamic, and evolutionary.

2. Panel (b) shows the dynamic process. Equilibrium of endogenous variable Y is given by the trajectory CE, for a given value of the exogenous variable X. The exogenous variable increases at time t', which implies a new dynamic equilibrium, given by trajectory DF; then point C′ is out of equilibrium and, because the equilibrium is stable, it will move to the new equilibrium along the trajectory C′D′, which is called the *transition dynamics*.
3. Panel (c) represents an evolutionary process. For a given value of the exogenous variable X_1, the "dynamic equilibrium" EE′ (which is actually a mirage, for it is only temporal) only operates until the threshold value of Y^*, at period T^*, is reached; beyond this period, the "dynamic equilibrium" breaks down. A new set of relations will appear, which is called the switching regime. Assuming that a new variable X_2 appears (an innovation at period T^*), and has a given value, the new trajectory of Y is represented by the curve FF′.

Mechanical processes and evolutionary processes imply different concepts of time: *mechanical time (t)* and *historical time (T)* (Georgescu-Roegen, 1971). This distinction can also be seen in figure 1.3. It is clear that in panels (a) and (b), the mechanical process allows the values or trajectories of the endogenous variable to go back if the exogenous variable goes back to its previous value, as if nothing had happened. There is no history. These cases use time t in the horizontal axis. In the evolutionary process, panel (c), it is clear that there cannot be a way back, for the economic process is going through qualitative changes. Time moves in only one direction; it is irreversible; it is historical time, in which past, present, and future are qualitatively different. This case uses time T in the horizontal axis.

It should be clear from the definition of equilibrium that it has no normative implications. Equilibrium could be socially desirable or undesirable. It does not show social optimality in any sense. Equilibrium does not mean absence of social conflict either; it does not imply social harmony. Equilibrium is just the aggregate outcome of social interactions, in which no social actor has the power and the will to change the situation. Economic equilibrium means social equilibrium as well, as it is the result of social interactions.

How does the alpha-beta method operate in mechanical and evolutionary processes? Mechanical processes present no problem, as indicated above. The evolutionary process presents some problems because it implies the existence of a threshold value of the endogenous value (Y^*) at which the economic process goes through regime switching or a breakdown of relationships. However, the threshold value Y^* is unobservable. Thus, the critical statistical test of an evolutionary model is about the existence of

breakdown of the temporary dynamic equilibrium. Now notice the following problem. If the breakdown has not happened, the theoretical model can be saved by the argument that the threshold value Y^* (unobservable) has not been reached yet; if it has happened, the theory is also saved because that is what it predicted. The theory would thus become immortal. The problem is that we cannot know what is going to happen in the future. Science can only explain the past, not the future.

However, the evolutionary model represented in figure 1.3 panel (c) is falsifiable because the trajectory EE′ is observable and may be either increasing or declining over time. If the observed trajectory EE′ is declining, then the evolutionary model will be rejected; if it is increasing, then the evolutionary model will be accepted, although provisionally, as indicated by the rule of acceptance of any theory.

On the concept of *runs*, short-run and long-run analyses correspond to logical categories in the economic process. We may assume that the longer the unity of time of repetition in the economic process, the more changes will take place in the process. Hence, we call short run when most variables are exogenous and long run when fewer variables are exogenous. The assumption is that as social actors face changes in the exogenous variables, they will seek to make adjustments, but adjustments take time; hence, more things will change in society in the long run than in the short run. Therefore, there will be more endogenous variables and less exogenous variables in the long run than in the short run. We can say that short-run and long-run categories refer to *logical time*.

This book will consider three types of runs that correspond to each of the types of economic process shown above. Short-run analysis will refer to the static economic process; long run to dynamic economic process; and very long run to evolutionary economic process. In the short-run analysis, the productive capacity of society is exogenously given; in the long-run analysis, the productive capacity is endogenous, as capital and labor grow endogenously, and technology grows exogenously, but the interactions between the economic process and the biophysical environment are ignored; in the very long run analysis, these interactions are considered essential to the economic growth process.

Although short run and long run are logical categories, they have some correspondence with chronological time. Thus, the short run can correspond to shorter periods (say, monthly or quarterly) compared to those of the long run (yearly), and the very long run (decades).

An Economic Theory Is a Family of Models

An economic theory cannot be falsified if it contains alpha propositions that are too general (in the domain of human intuitions only, such as "people

act guided by their own interest") because beta propositions could hardly be derived from them. If this is the case, the alpha propositions will constitute the *primary assumptions* of the theory, and now additional assumptions, called *auxiliary assumptions*, will be necessary to make the theory situational (with a given social context) and then operational, that is, able to derive beta propositions. The set of auxiliary assumptions gives rise to a *model of the theory*. A theoretical model then includes two subsets of assumptions, such that the auxiliary assumptions cannot contradict the primary assumptions. A model is a logical system.

An economic theory is then a family of models. The set of alpha propositions constitutes the common element or the core of the family. Beta propositions will now be derived from models. The theory is then subject to the process of falsification through its models. This falsification requires that the number of models be finite. Therefore, and more precisely, an economic theory is a family of *finite* number of models.

Figure 1.4 illustrates the algorithm of the alpha-beta method in economics. Given a theory α, a set of consistent auxiliary assumptions A' is included to construct the model α', from which the set of empirical predictions β' is derived, which is then tested out against the set of empirical data $\mathbf{b}$. If model α' does not fail the statistical test, then the theory can be accepted; if it fails, then model α' is rejected, but the theory is not. Using auxiliary assumption A'', another model (α'') is constructed and submitted to the falsification

$$\alpha\text{–}(A') \rightarrow \alpha' \Rightarrow \beta' \rightarrow [\beta' \approx \mathbf{b}]$$

If $\beta' = \mathbf{b}$, then α' explains reality and so does α

If $\beta' \neq \mathbf{b}$, then α' fails to explain reality; then

$$\alpha\text{–}(A'') \rightarrow \alpha'' \Rightarrow \beta'' \rightarrow [\beta'' \approx \mathbf{b}]$$

If $\beta'' = \mathbf{b}$, then α'' explains reality and so does α

If $\beta'' \neq \mathbf{b}$, then α'' fails to explain reality; then algorithm is followed

$$\alpha\text{–}(A^n) \rightarrow \alpha^n \Rightarrow \beta^n \rightarrow [\beta^n \approx \mathbf{b}]$$

If $\beta^n \neq \mathbf{b}$, then α^n fails to explain, so does α. Then, a new theory is constructed and the algorithm continues.

Figure 1.4 Alpha-beta method in economics.

process; if it does not fail, then the theory is accepted; if it fails, the theory does not fail, and then another model is constructed, and so on, until model α^n. This algorithm requires, as the definition put forward above says, that the number of models be finite, that is, it requires that the theory can generate only a limited number of possible auxiliary assumptions. If all models of the theory fail, then the theory fails, and a new theory will be needed.

What are the auxiliary assumptions about? Basically, auxiliary assumptions refer to the particular context in which social actors operate. In the study of the capitalist society, for example, consider an economic theory that assumes the market system and the democratic system as the basic institutions. Then the following possible social contexts can be established:

(a) On the market system: assume the particular degree of market power (perfect competition—absence of market power—monopoly, or oligopolistic competition);
(b) On the democratic system: assume the particular political power structure (highly participatory, representative, or authoritarian);
(c) On the nature of the economic process: static, dynamic, or evolutionary.

A combination that results from the selection of one element from each category will determine a particular model of this economic theory of capitalism. The total number of possible combinations is clearly finite.

It should be noted that the use of models does not constitute protective belts of theories to avoid falsification, as some epistemologists have suggested (the so-called Duhem-Quine problem). On the contrary, the use of finite number of models makes an economic theory falsifiable. A protective belt of an economic theory could result from introducing unobservable factors into the models, such as probabilities, expectations, propensities, preferences. Hence, a model that assumes expectations will become nonfalsifiable: Whenever the model fails, it can be saved by arguing that expectations of people have changed. Expectations may be introduced into the model but as dependent upon an observable variable, as a structural equation, which in the reduced form will lead to refutable empirical predictions or beta propositions.

Partial Theories Must Constitute a Unified Theory

The purpose of process analysis is to construct an abstract world that resembles the social reality under study. This representation can refer to parts of the social world or to its unity, which gives rise to the construction of partial

theories and the unified theory. Each level of analysis can be carried out through the use of models, which are called partial equilibrium models and general equilibrium models.

Good or valid partial theories thus constitute fragmentary knowledge of the real world and are not necessarily conducive to the unity of knowledge, the unified theory. Only unified theories can then provide scientific knowledge. A valid unified theory explains the parts, taken separately, and also the whole reality under study.

In sum, the study of capitalism that we are going to undertake in this book will utilize the alpha-beta method, which is a particular method of Popperian epistemology, as shown in this chapter. We have established here the rules of scientific knowledge that we are going to apply for rejecting or accepting economic theories that seek to explain production and distribution under capitalism. These rules were summarized in figure 1.4. The alpha-beta method will be seen in action in the rest of the book. The empirical regularities of the capitalist system (represented by letter **b**) are the facts that a good economic theory should be able to explain. The basic facts to be used in the task of falsifying economic theories along this book are presented in the next chapter.

CHAPTER 2

The Capitalist System: Empirical Regularities

This book seeks to explain the economic process in the capitalist system, and therefore we need to know what are the facts to be explained. This chapter presents a set of empirical regularities about production and distribution in capitalist countries, which has been taken from the international literature.

Scope of the Book: Capitalist Countries, 1950–2010

The theoretical definition of capitalist societies was given in the previous chapter. The empirical counterpart of the capitalist system will refer to those independent countries that have been operating since post–World War II, or most of the time, under the institutional norms of capitalism in which the following hold good: (a) private property of physical capital is more predominant than state property; (b) the market system is more predominant than nonmarket forms of exchange of goods between individuals; (c) democracy is the predominant form of governance, although it includes different degrees of people participation in public policies.

According to this empirical definition, and using World Bank statistics, there are 174 countries in the world that may be classified as capitalists during all or most of the period 1950–2010. They will be divided into two groups: the *First World* and the *Third World*. The First World consists of the richest 23 countries, according to World Bank estimates. They are: Australia, Canada, Japan, New Zealand, United States, and the 18 countries of Western Europe (Austria, Belgium, Denmark, Finland, France, Germany, Greece,

Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and United Kingdom). The rest of the 151 countries constitute the Third World.

Just for the sake of completeness, we should say something about non-capitalist countries. Socialist or communist countries are characterized by the predominance of state property of physical capital over private property, the predominance of state planning over markets, and authoritarian political regimes over democracy in all or most of the period 1950–2010. This category, sometimes called the *Second World*, consists of 33 countries. They can also be separated into two groups. The first group consists of countries "in transition" to capitalism due to the introduction of market reforms since the 1990s, which include the countries of Eastern Europe, the Balkans, and the ex-Soviet Union. In the second group, those that have remained as communists in most of the period of analysis, we find only five countries, which include China, Cuba, Laos, North Korea, and Vietnam.

This classification intends to be stylized, ignoring some details. For example, some capitalist countries of today were under communist regime for short periods. However, considering countries that were under a communist regime for 10 to 17 years, the longest periods recorded, the list includes only six countries: Angola, Benin, Congo, and Mozambique in Africa, and Afghanistan and Cambodia in Asia.

The scope of this book is the study of the capitalist system only: production and distribution in the First World and the Third World, as defined above. The period of analysis will be limited mostly to the period 1950–2010. This selection is due to the constraint in availability of statistical information.

Surely, the book is not about production and distribution in the entire world economy. Therefore, it should be clear from the outset that the Second World countries, as defined above, are not part of the analysis of the book.

In particular, the reader should keep in mind that China is not part of the analysis presented in the book. Given its large population size and its rapid economic growth in the last decades, China has changed the level and the structure of the world economy. However, China is not a capitalist country even today and cannot be introduced into a theoretical system that assumes the capitalist form of production and distribution. China's economic growth has increased the size of world output and international trade, but it has not changed the rules under which capitalist countries operate. So China can safely be ignored in the analysis of capitalism over the past six decades. The empirical data on production and distribution in the capitalist countries presented in the book include implicitly the China effect, although the size of this change will not be singled out.

A Brief History of Capitalism and Colonialism

The origins of capitalism are usually traced to the industrial revolution that took place at the end of the eighteenth century and beginning of the nineteenth century in Western Europe. If we take 1800 as the initial point, then capitalism has operated for nearly two centuries now.

How have capitalist countries come about? The well-known proposition of Karl Marx says that societies tend to march along the same path in their history. "Marx saw all countries as following the same path from slavery to feudalism to capitalism, and finally to socialism and communism" (Nolan, 2006, p. 353).

Consider a simplified trajectory in which societies start with a primitive or natural form of social production and distribution; moving then to more complex forms, such as the feudal form, characterized by agrarian societies, in which landlords concentrate the property of land and thus use landless workers for production under bondage systems; the capitalist form of production, characterized by industrial societies, in which capitalists concentrate the property of physical capital and hire free workers in the labor market; and the communist society.

To these types of human societies, European colonialism will now be introduced. The evolution that today's world countries have experienced could be summarized in a stylized manner using those categories, since the discovery of America. This is shown in table 2.1. Current countries of the world are grouped into eight categories. The first four groups correspond to countries that have functioned as capitalist societies in most of the period of analysis (1950–2010); group (5) refers to what is called countries "in transition" from communism to capitalism, since the 1990s only; finally, the last three groups refer to the five countries that most of the time in the period of analysis have functioned as communist countries.

Some comments on table 2.1 are in order. Colonialism will refer to the domination of regions of the world by Europeans. The domination of a society upon others within the same region, either within Europe or within Asia, is not part of this definition. The idea is that the domination of the New World societies, which were very different in technology and culture, by Europeans is what becomes significant in the postcolonial period as historical legacy. A distinction is also made between colonialism and settlement. Colonialism implies domination by Europeans of regions significantly populated and during a significant period; settlement implies domination of regions that were mostly empty lands, with low population density.

In table 2.1, group (1) includes Western Europe. Some of these countries were colonial powers (England, Spain, Portugal, France, Germany, Holland,

Table 2.1 Evolutions of countries to capitalism, 1500–2010

1.	Western Europe: *Naturalism—Feudalism—Capitalism*
2.	Never colonized by Europeans (Afghanistan, Ethiopia, Iran, Israel, Japan, Singapore, South Korea, Taiwan, Thailand, and Turkey): *Naturalism—Feudalism—Capitalism*
3.	Settlements by Europeans: (a) United States, Canada, Australia, and New Zealand; (b) Argentina, Chile, Costa Rica, and Uruguay: *Settlement—Capitalism*
4.	Colonized by Europeans: Most of Africa, Asia, and Latin America: *Naturalism—Colonialism—Capitalism*
5.	Eastern Europe (Ex-Soviet Union, Eastern Europe, Balkans, Mongolia): *Naturalism—Feudalism—Communism—Capitalism*
6.	China: *Naturalism—Feudalism—Communism*
7.	Cuba, Laos, Vietnam: *Naturalism—Colonialism—Capitalism—Communism*
8.	North Korea: *Naturalism—Colonialism—Communism*

Source: Elaborated by the author from Dalziel (2006) and Wesseling (2004).

Italy, and Belgium). Group (2) includes countries that were never European colonies. Group (3) refers to European settlements. In the New World, the United States and Canada were not colonies, but rather settler territories, a distinction made by historians:

> North America was a different matter. Here there were no Amerindian settlements of sufficient size to be worth conquering and consequently no firm bases for further exploration. The conquistador De Soto found few Indians and no gold. (McEvedy 1972, p. 20)

The same statement could be applied to Australia and New Zealand and also to Argentina, Chile, Costa Rica, and Uruguay in Latin America. Group (4) includes countries that were colonized by Europeans and thus have a strong colonial legacy.

As can be seen in table 2.1, the predictions of Marxian theory are refuted by the historical facts: the sequence *naturalism-feudalism-capitalism-communism* has not been observed in any country of the world. Most countries that adopted communism were feudal systems before. It seems that Marx assumed the history of Western Europe, which is *naturalism-feudalism-capitalism*, as the typical evolution and extrapolated it to the rest of the world. In most Third World countries of today, capitalism has emerged

from colonialism, as indicated by group (4). European colonialism has certainly played a significant role in shaping the observed historical path of countries.

From the 151 countries classified as Third World, only 13 have weak or no European colonial legacy. These include the nine countries of group (2), excluding Japan. Afghanistan, Iran, Ethiopia, Thailand, and Turkey were never subject to European colonial domination; other three Asian countries have not been colonized by the Europeans, but they have been by the Chinese and Japanese, although for a short period of time only (less than 50 years): Taiwan, South Korea, and Singapore (Wesseling 2004). Israel, which was founded in 1948, is also included in this group. If we add the four countries of Latin America that were mostly settlements, group (3b), then 13 Third World countries of today have none or a weak colonial legacy. Therefore, the large majority of Third World countries, 138 out of 151, have a strong colonial legacy.

The scope of this book is the study of the capitalist system. Groups (1) through (4) in table 2.1 are then the relevant categories for this study, 174 countries in total. It should be clear that the criterion to define these groups was qualitative: the European colonial history of capitalist countries. The current capitalist system has therefore the following structure:

1. *First World*: Groups (1) and (3a) plus Japan, 23 countries,
2. *Third World with Weak Colonial Legacy*: Groups (2) and (3b), 13 countries, and
3. *Third World with Strong Colonial Legacy*: Group (4), 138 countries.

We can now apply a quantitative criterion to establish the following relationship: countries that were never under European colonialism are the rich ones (First World); countries that have a strong legacy of European colonialism are the poorest countries; and countries that have a weak or no colonial legacy are middle-income countries, as shown in table 2.2.

Empirical Regularities on Production, Distribution, and the Environment

The aggregate behavior of the three groups of capitalist countries established earlier is now presented as a set of empirical regularities. These regularities will basically refer to production and distribution variables. They are statistical regularities, that is, they are valid in the vast majority of countries, or in the average, but not necessarily in every country. The remainder of the book will intend to explain these empirical regularities.

Table 2.2 Income level and income inequality in capitalist countries

Group	*Average income US$, PPP*		*Income inequality Average Gini index*	
	1980	*2008*	*1950–1970*	*1971–2008*
1. First World/*Epsilon*	9,508	38,563	0.36	0.33
	(1,977)	(9,570)	(0.04)	(0.04)
	23	23	15	22
2. Third World with weak colonial legacy/*Omega*	3,779	19,473	0.39	0.36
	(2,052)	(13,312)	(0.09)	(0.07)
	9	9	5	7
3. Third World with strong colonial legacy/*Sigma*	2,360	6,022	0.47	0.49
	(3,662)	(6,544)	(0.09)	(0.07)
	65	72	30	32

Notes: Figures in parenthesis indicate standard deviation and those in the third row number of countries in the sample. Total number of countries in each category is 23, 13, and 138.
Average income: It is Gross National Income per capita, adjusted by purchasing power parity (PPP) to reflect differences in the purchasing power of the American dollar in different countries relative to the United States. Samples of First World countries include: Australia, Canada, Japan, New Zealand, the United States, and the 18 countries from Western Europe; Third World countries with weak colonial legacy include: Argentina, Costa Rica, Iran, Israel, Singapore, South Korea, Thailand, Turkey, Uruguay; and Third World countries with strong colonial legacy include a total of 73 in both periods.
Gini Index: Considers only comparable data across countries: the distribution of net income per capita calculated from national household surveys. Third World countries with weak colonial legacy include Costa Rica, Israel, Singapore, South Korea, Taiwan, Thailand, and Turkey. In order to have sufficient sample size, it was necessary to define 1950–1970 as the initial period.

Sources: Elaborated by the author from the following sources:
Average income: Calculated from World Bank—Data. http://data.worldbank.org/indicator/NY.GNP.PCAP.PP.CD/countries. Last date of revision: Not indicated. Last date accessed: 19/10/2011.
Gini Index: Calculated from the database of Branko Milanovic (2010), Web Page at World Bank, "All the Ginis Dataset": http://go.worldbank.org/9VCQW66LA0. Last date accessed: 19/10/2011.

The production and distribution regularities within and between the First World and the Third World in the period 1950–2010, for which statistical data are available, can be summarized as in the following sections.

Fact 1

Existence and persistence of unemployment in the First World. Unemployment is generally rife in First World countries. Unemployment rates may vary across countries and across periods within a country, but these rates are always positive. The US economy showed average annual rates of unemployment that varied from 3 percent to 10 percent in the period from 1960 to 2005, whereas in Western Europe these rates were 2 percent and 11 percent in the same time period (Blanchard 2009, Table 1.5). The new feature of the ex-socialist economies in transition to capitalism is the fact that unemployment has appeared as a new social problem, that is, capitalism and unemployment arrived together. Unemployment is the disease of capitalism, as stated by the historian John Garraty (1978).

Fact 2

Existence and persistence of unemployment and underemployment in the Third World. Three groups of workers can be distinguished in the Third World. The first is composed of those workers employed in firms who earn market wages. The second is the unemployed. Unemployment rates are generally positive in Third World countries. These rates also show variations by countries and by periods, and with rates in the range that is similar to those of the First World. The third is the group of workers who are self-employed in small businesses and small farms (with no hired labor).

Self-employment rates are more significant in the Third World than in the First World. There are estimates for urban areas that show this difference. According to ILO statistics, in the period 1980–2000, the rate of self-employment in the First World was 12 percent on average (over a sample of 17 countries), whereas in the Third World it was 40 percent (over a sample of 58 countries) (ILO 2002, Annex 2, pp. 62–64). If the rural areas were included, this gap would be even higher, because self-employment rates are surely much higher in rural than urban areas and this gap is probably higher in the Third World.

Empirical studies show that the average income of the self-employed in the Third World is smaller than the average wage prevailing in the labor market, for a given level of skills (Figueroa 2010; Rodriguez 2013; Telles 1993). The situation of the self-employed is therefore involuntary and undesirable, as in the case of the unemployed, because these workers would prefer to have wage employment. This group is called the *underemployed*.

The excess labor supply takes the form of unemployment in the First World, whereas in the Third World it is composed of unemployment and

underemployment, the latter being the more significant. Therefore, the common practice of using the unemployment rate as the criterion for making international comparisons about the excess labor supply is unwarranted. Unemployment figures underestimate the magnitude of the total excess labor supply in the Third World.

Fact 3

Existence and persistence of income gaps between ethnic groups in the Third World. Most Third World countries are multiethnic. Empirical studies on the income gaps between ethnic groups in countries of Latin America, Africa, and Asia have found the existence and persistence of these gaps (Darity & Nembhard 2000; Figueroa 2010; Hall & Patrinos 2005, 2012; Psacharopoulos & Patrinos 1994; Silva 2001; Stewart 2001; Telles 1993). These empirical studies also show that gaps are systematic, that is, ethnic groups maintain their positions in the income pyramid of their respective countries. In particular, those ethnic groups that are descendants of aboriginal populations or slave populations—the legacy of colonial systems—have the lowest income levels.

Fact 4

In the short run, nominal and real variables are correlated in both the First World and the Third World. In the workings of capitalist countries, empirical studies on the relationships between changes in the nominal variables under the control of governments (money supply, exchange rate, interest rate) and real variables (total output, real wages, and employment) in the short run have found that they are correlated. For instance, monetary aggregates and total output tend to move in the same direction in the US economy, based on quarterly data from 1959 to 1996 (Barro 1997, Table 18.1, p. 705). These empirical relations have also been found in a sample of 10 countries of the First World and 15 countries of the Third World (from sub-Saharan Africa, Latin America, Asia, and North Africa) and for the period 1970–1997 (Rand & Tarp 2002).

Fact 5

In the long run, real wage rates and output per worker are positively correlated in both the First World and the Third World. When output per worker grows for long periods, real wages also tend to increase. For example, in the US economy, in the period 1950–1974, output per worker grew at the annual

rate of 1.9 percent per year, while real wage grew at a similar rate (Barro 1997, Chapter 6, pp. 225–226). In the Third World, a sample of 12 Latin American countries showed positive growth rates in both output per capita and real wages in all but three countries in the period 1980–1999 (ILO 2000, Table 9A). There is no such thing as workers immiserizing growth process.

Fact 6

In the long run, the income-level gap between the First World and the Third World is persistent over time. Table 2.2 shows that the average income level of First World countries is higher than that in the Third World. This does no more than checking our definition. What is significant is that this gap has shown persistence over time. The ranking of the three groups of capitalist countries, as were defined earlier, remains unchanged for 1980 and 2008 as well. Therefore, table 2.2 shows a significant relationship between the type of colonial legacy of capitalist countries and the income levels achieved for a large sample of capitalist countries. There exists a statistical correspondence between colonial history and average income level among capitalist countries. Western European countries and countries that were European settlements are high-income countries today; countries that were never European colonies or have a weak colonial legacy tend to be middle- income countries, whereas countries with strong colonial legacy tend to be low-income countries.

There are some exceptions, however. Argentina, Chile, Costa Rica, and Uruguay should be high-income countries (as Canada, Australia, and the United States), but they are middle- income countries. In fact, Argentina was a high-income country in the nineteenth century and at the beginning of the twentieth. According to historian Angus Maddison (1995), while Japan is the only case of success, Argentina is the only case of involution. Afghanistan and Ethiopia should be middle-income countries (as Taiwan and South Korea), but they are low-income countries.

Table 2.2 also shows that between 1980 and 2005 the income gap has increased from 4.0 times to 6.4 times between the First World and the Third World with strong colonial legacy. The gap between the First World and the small group of countries of the Third World with weak colonial legacy has declined from 2.5 to 2.0. In the very long run, the estimates made by economic historian Angus Maddison indicate a drastic *increase* in the income gap. The income per capita of the richest countries compared to the poorest was 3:1 in 1820, which jumped to 18:1 in 2000 (cited in Galor 2011, Figure 1.1, p. 2).

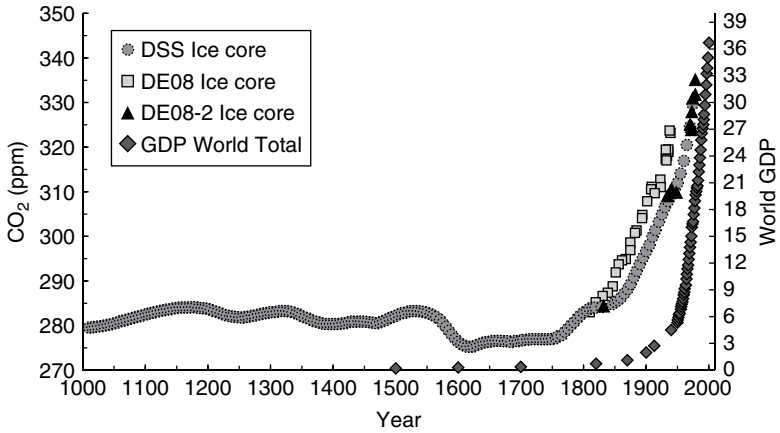


Figure 2.1 CO_2 (1000–1998) and World Total GDP (1500–2000).

Notes: World Total GDP estimated by Maddison (2003) and the yearly data taken from http://www.ggdc.net/maddison/historical_statistics/horizontal-file_03-2007.xls. They are measured in 10^{12} units at 1990 International Geary-Khamis dollars. CO_2 measured as part per million from air extracted from three Antarctic ice cores: DE08 to provide a high resolution record of atmospheric methane over the industrial period; DE08–2, to find the effects of diffusion on the composition and age of the enclosed air and to better establish the link between it and the atmosphere; DSS, the new Australian deep ice core, extends the CO_2 record back into the late Holocene, taken from Etheridge et al. (1998): http://cdiac.ornl.gov/trends/co2/lawdome.html.

Source: Maddison (2003) and Etheridge et al. (1998).

Fact 7

In the long run, the degree of within-country inequality in the Third World is, on average, higher than that in the First World, and this difference is persistent over time. Table 2.2 also shows estimates of the differences in the average degree of income inequality among households in the three groups of capitalist countries. The degree of income inequality is measured by the Gini index. (This index varies between zero and one. The value of zero indicates perfect equality in the distribution of income, whereas the value of one indicates perfect inequality, all income is concentrated in the hands of one household). Certainly, the empirical Gini index can hardly reach the value of zero or one. On the one hand, the observed empirical values calculated in the international literature hardly reach values beyond 0.70; on the other hand, the lowest values observed rarely go below 0.20. Therefore, the use of the terms "low-inequality" and "high-inequality" societies in this book will be in reference to these *actual* extreme values.

As seen in table 2.2, the within-country inequality is, on average, lower in the First World compared to what it is in the Third World with strong colonial legacy, while the value in the Third World with weak colonial legacy lies in between. It is clear that the First World is not only wealthier than the Third World, but also more egalitarian. This empirical correspondence between income levels and degrees of equality constitutes another empirical law of the capitalist system. Over time, we can see increasing inequality in the Third World with strong legacy, and decreasing in the other groups; but the variations are small, and may just reflect changes in the composition of the samples. What is clear is that the order of inequality among country groups tends to remain unchanged.

In the international literature, empirical studies by single countries have shown that the Gini index for both the First World and Third World countries shows significant viscosity in the period 1950–1995 and the same order shown in table 2.2 (e.g., Atkinson 1996; Deininger & Squire 1996; Li, Squire, & Zou 1998). These estimates refer to household incomes. Household income data are the most common source in income inequality studies, but they capture mostly labor income (wages, salaries, and income of self-employment), leaving aside profits and rents, which mostly accrue to the very wealthy, that is, household income inequality tends to underestimate the real inequality in society. When empirical data on national income distribution are available, as will be explained next, data still support Fact 7.

Fact 8

In the very long run, the process of economic growth is accompanied by the degradation of the biophysical environment. In contrast to the previous regularities, Fact 8 refers to the economic process in the very long run, the last two centuries. From the biophysical environmental point of view, the world economy is a closed system. The observed environmental degradation of the last 200 years will then have to be related to the world output, not only to that of the capitalist system. Around year 2,000, the capitalist system produced 83 percent of the world output and concentrated 71 percent of the world population (World Bank 2001, Table 1, pp. 274–275); hence, world output in our period of analysis (1950–2010) reflects mostly output produced in the capitalist system. The same source shows that China produced 11 percent of world output with a share of 21 percent of world population; moreover, the effect of China's growth on environmental degradation would be reflected in the changes of the last decades only.

Economic historian Angus Maddison has made calculations about the economic growth experience of the world economy as a whole during the

period 1820–2001: total output grew at the average rate of nearly 1 percent per year in the period 1820–1950, but at the rate of 4 percent in the period 1950–2001 (Maddison 2003, pp. 256–257). This global economic growth performance is highly correlated with degradation of the environment, which includes, among other indicators, pollution. Figure 2.1 presents this empirical relation, where pollution is measured by the concentration of carbon dioxide (CO_2) in the atmosphere and global output by GDP. The trajectories of these two variables are positively correlated between the years 1500 and 2000: they were both rising rapidly in the period 1950–2000, and both remained almost stagnant before this period. Studies also show a positive correlation between pollution and global warming; thus, Robert Stavins has summarized graphically the existence of this global positive correlation for the period 1880–2010 (Stavins 2011, Figure 4, p. 170). Then the theoretical question arises: How does one explain these observed relationships between output growth, environmental degradation, and global warming?

The Need for a Unified Theory to Explain the Empirical Regularities

Does economics have a theory that is able to explain the eight empirical regularities? If it does, such a theory will have achieved unity of knowledge in explaining why capitalism functions in this manner.

Neoclassical theory and Keynesian theory constitute the standard economics of today. These two theories complement each other. Keynesian theory is utilized in the short-run analysis (inflation, unemployment, business cycles) and the neoclassical theory mostly in the long-run analysis of economic growth. The analytical reason for the complementarity is that Keynesian theory becomes neoclassical theory in the long-run analysis. Although there is still a debate between these two theories over their capacity to explain the short run, neoclassical economics has become dominant for the long-run analysis. As will be shown in chapter 3, standard economics is unable to explain all the empirical regularities established in this chapter. Classical economic theory does not either. Consequently, a unified theory of capitalism is needed.

As displayed in table 2.2, income level and degree of inequality of capitalist countries are not independent of their colonial history. This is one of the basic empirical regularities of capitalism, after 200 years of functioning. Certainly, this regularity is an economic law and needs a theoretical explanation. This theory will be provided in this book. For this purpose, the three categories of capitalist countries established earlier will be represented by three abstract capitalist societies:

1. First World: *Epsilon society*
2. Third World with weak colonial legacy: *Omega society*
3. Third World with strong colonial legacy: *Sigma society.*

These abstract societies constitute partial theories of the capitalist system, which intend to explain the three types of real capitalism, taken separately; then a unified theory will also be presented to explain the functioning of the capitalist system, taken as a whole.

The challenge is enormous. The unified theory of capitalism we need must explain the economic process in the short run and in the long run. This theory must also explain the economic process in the very long run, when degradation of the biophysical environment is taken into account as part of the economic process. The requirement is that the needed economic theory must generate unity of knowledge. This calls for unity of knowledge in economics. Fragmentary knowledge on each part of the reality, or valid partial theories that are not conducive to a unified theory, will not help.

Old and new economic theories are then presented to explain capitalism. The rules of scientific knowledge used to reject and accept theories throughout the book correspond to those of the Popperian epistemology, which was developed in the previous chapter in the form of the alpha-beta method. Standard economics and classical economics have a significant heuristic value in the construction of new theories. Surely, one can make progress more easily if one is standing on the shoulders of giants, and giants were those who constructed and developed the science of economics up to now. Therefore, the new theory presented in this book will not be fully understood unless the foundations of received economics are understood. These foundations are presented in the next chapter.

CHAPTER 3

Standard and Classical Economics

Several theories that seek to explain production and distribution in the capitalist society coexist in economics. Neoclassical, Keynesian, and classical theories constitute the most important ones. Are these theories able to explain the eight empirical regularities of capitalism presented in chapter 2? This chapter presents, first, the foundations of these theories, then follows short-run models and their empirical predictions, which are finally confronted against the relevant empirical regularities.

Following the alpha-beta method, these theories are seen as abstract societies. Therefore, the failure of a theory in explaining one out of the eight empirical regularities is sufficient to reject it.

The Neoclassical Society

The construction of the abstract neoclassical society presented in this section is inspired by the works of Adam Smith (1776) and Leon Walras (1883). The primary assumptions of neoclassical theory are the following. As regards institutional context, social interactions operate under the rules of the market exchange and democracy. In this society, people participating in the economic process are endowed with quantities of economic and political assets. Assets are goods that provide a flow of incomes. As regards economic rationality, people act motivated by self-interest.

Therefore, people exchange goods in the form of market exchange, which is subject to the following rules:

(a) It is voluntary.
(b) It is based on the motivation of self-interest, which implies that people seek to take advantage of every exchange possibility to make economic gains for their own benefit.
(c) It is constrained by the individual's endowments of economic assets only, which means freedom of exchange, that is, absence of social or legal restrictions. If legal restrictions apply, then market exchange is regulated.

An implication of market exchange is that it is impersonal, which leads to exchange balance in every transaction. Consequently, the law of one price for buyers and sellers will prevail in each market.

In contrast, nonmarket exchange of goods may take the form of either command or reciprocity. Command clearly implies absence of rules (a), (b), and (c), as in the case of forced labor exchange (slavery systems). Reciprocity includes rules (a) and (b) only, for the exchange of goods is part of exchange of favors; hence, social assets and social norms play a significant role. People still act motivated by self-interest, but personal relations are now essential for the exchange of goods. Thus, the exchange balance need not hold in the short run; it will be attained in the long run; hence, the law of one price need not apply in the short run. This form of exchange is restricted to social networks, such as extended families, social groups, and communities, particularly traditional peasant communities, where the market system is not well developed (Figueroa 2009).

In the neoclassical society, all markets are Walrasian. A Walrasian market is a form of market exchange in which individuals are able to buy or sell the good under exchange in the quantities they desire at the prevailing market price, that is, no one is left without realizing the desire transaction. If this situation were not attained (if there were excess demand or excess supply), then the price of the good would have to change, until the market was cleared.

A Walrasian market assumes fully flexible prices. A market price plays the role of a rationing mechanism: if the quantity demanded of a good is higher than the quantity supplied, then the market price will increase; if it is the other way round, then the market price will fall. The assumption is that the market system is self-regulated, that is, it operates *as if* it were a big computer solving a system of equations.

The alpha propositions of the neoclassical theory (N) are the following:

α*(N). (1) Institutional context*: (a) Rules: People participating in the economic process are endowed with economic assets, which are subject to private property rights; people exchange goods subject to the norms of market exchange. The market system is composed of Walrasian markets. The political regimen is democratic. (b) Organizations: households, firms, and the government.

α*(N). (2) Initial conditions*: Factor endowments of society (capital per worker) are such that the productivity of total labor is high enough for firms to pay wages and generate profits.

α*(N). (3) Economic rationality of agents*: Individuals act motivated by self-interest.

A Neoclassical Model: The Short Run

In order to make neoclassical theory falsifiable, a neoclassical model must be constructed, which can be seen as constructing a particular stage set in which social actors play their roles. The following auxiliary assumptions are introduced for that purpose:

- There are two social groups: capitalists who are endowed with stocks of physical capital and human capital, and workers with stocks of human capital and cash balances.
- Government behavior consists of supplying money.
- Three markets—all Walrasian—constitute the market system: labor, commodity, and money. Workers are endowed with the same human capital. Hence there is only one labor market. Only one good is produced in society, called good B. Workers and capitalist firms exchange labor services and goods in the labor and commodity markets; money supplied by the government is used as the means of payment and is exogenously determined. Wages are paid at the end of the production period; consequently, workers need to hold cash balances for transaction purposes. There is no market for renting capital services.
- Capitalists and workers interact and compete in the market to obtain their objectives of self-interest, but none has the power to set prices (market structure of perfect competition); they all face costless information on market prices and technology. Labor productivity levels vary across firms.
- The economic process is static; it is closed as well, for market exchange with other societies is ignored.

"Short run" and "long run" correspond to logical categories and to *logical time*. Short-run analysis assumes that the productive capacity of society has a given value, whereas long run assumes that this capacity changes. However, logical time has some correspondence with chronological time. Thus, short run must correspond to an economic process that is repeated in short chronological periods (say, monthly or quarterly), in which variations in the productive capacity will be small and thus may safely be ignored. Hence, there will be no capital accumulation in the short-run model. Changes in the productive capacity will be analyzed through "long run" equilibrium models.

With these auxiliary assumptions, we have a neoclassical model, which becomes operational, for beta propositions can be derived from this theoretical model. These beta propositions can then be confronted against the short-run empirical regularities of capitalism.

Behavior of Workers

In the analysis of the economic behavior of social actors, consider first the social group of workers. Workers seek to maximize the quantity of good B for consumption, subject to their budget constraint and their real cash balance constraint. Workers need to hold a quantity of real cash balance for transaction and precautionary purposes. The stock of money to be held as cash balances is a requirement to make transactions in the market, according to the practices of payments. The real cash balance depends upon the real income: the higher the workers' income, the higher the real cash balance required. But, this requirement is not proportional: the model will assume that double real income requires less than double real cash balance.

Therefore, for given total real incomes, there is an optimum stock of money that workers must hold. If the cash balance held is higher than the required, workers will seek to run down the stock by spending the excess in the purchase of good B; if the cash balance is smaller, they will seek to build up the stock by buying a smaller quantity of the commodity. The adjustment mechanism to get the optimum cash balance operates via changes in the quantities of good B bought. Assume this adjustment in stocks is made in one period. Once this adjustment is made, workers will buy the quantity of good B that their wages permit, the equilibrium quantity, period after period.

Workers exchange their labor for good B in the market. They are able to sell all their labor in the labor market at the prevailing wage rate and are also able to buy all the quantity of good B that they are willing to exchange because they exchange these goods in Walrasian markets.

Workers are price takers in both markets, that is, they take both the nominal wage rate (P_h) and the nominal price of good B (P_b) as given. These are the exogenous variables upon which workers cannot decide. Their endowments are also exogenously determined: labor (S_h) and nominal cash balance (S_m). The quantity demanded of good B (quantity D_b) and the real cash balance (D_m/P_b) are the endogenous variables upon which workers can decide.

The behavior of the individual worker i can be represented by the following system of structural equations:

$$P_h S_{hi} = P_b D_{bi} + (D_{mi} - S_{mi}) \qquad (3.1)$$
$$D_{mi}/P_b = L_i (D_{bi}),\ L_i' > 0,\ L_i'' < 0.$$

The first equation shows the worker's budget constraint; the second is the demand for real cash balances. The equilibrium condition is that the individual's cash balances are willingly held ($S_{mi} - D_{mi} = 0$). The endogenous variables include the quantities of good B bought and the real cash balance held willingly. As long as the exogenous variables remain fixed, workers will maintain the equilibrium values of the endogenous variables constant period after period. This situation represents a static equilibrium.

This equilibrium is (trivially) stable. Therefore, the method of comparative statics can be applied to derive beta propositions, that is, to determine the effect of exogenous variables upon endogenous variables.

An increase of the nominal price of good B (the price level) would have the effect of reducing the quantities of both good B bought and real cash willingly held. The reason is that real wage rate would fall and the quantity demanded of good B would also fall. At this lower level of real income, the required level of real cash balances would fall. But the rise in the price would lead precisely to a fall in the initial real cash balances. Then the new and old stock would tend to equalize. Of course, if the relation between real income and real cash balances was proportional, these values would indeed equalize; but the relation is not proportional, as workers would end up with a deficit of real cash balances. They would have to increase the stock of money, which implies buying a smaller quantity of good B in the period of adjustment. There would be a *real-cash balance effect* on the demand for good B. Once this adjustment is made, the new equilibrium implies lower real income and consumption of good B, together with a lower amount of real cash balance, period after period.

A rise in the nominal wage will have the effect of increasing the quantities of both good B bought and real cash balance held. This is because the real wage rate will now rise and then the effect will be just the inverse of that seen above, when the price level increased.

Last, an increase in the worker's money endowment will have no effect. The reason is that the increase in money endowment generates excess in his holdings of cash balances. The worker will seek to get rid of this additional quantity of money by buying an extra quantity of good B. Assume this adjustment is made in just one period; hence, in the adjustment period, the worker will buy an additional amount of good B. Once this adjustment is made, the equilibrium quantity of cash balances is restored and the worker will continue to consume the initial amount of good B, holding the initial nominal (and real) cash balance, period after period.

The real cash balance effect of a change in the price level or in the nominal wage rate upon the quantity of good B demanded operates only in the adjustment period. Over several periods, where the equilibrium values prevail, this effect will be small and may be ignored. A change in the money endowment of the individual worker has a similar effect as the real cash balance effect. The effect upon the quantity of good B demanded is positive but appears in the adjustment period only, that is, over several periods the effect is small and may be ignored. Hence, real cash balance effects can be safely ignored.

On the behavior of the individual worker, in sum, changes in the values of the exogenous variables have the effect of changing the values of the endogenous variables to a new equilibrium situation. The result is that the quantity demanded of good B coming from the individual worker depends directly upon the nominal wage rate and inversely upon the price level. If both nominal prices increase in the same proportion, which means maintaining the real wage rate fixed, the quantity demanded will remain unchanged; hence, the quantity demanded of good B will depend upon the real wage rate only. This relationship can be represented as the individual worker's demand function for good B as follows:

$$D_{bi} = f^i (P_b, P_b, S_{hi}) = F^i (w, S_{hi}), F^i_1>0, F^i_2>0, \tag{3.2}$$

subject to the constraints indicated in equation (3.1).

Behavior of Capitalists

Consider now the behavior of the social group of capitalists. Capitalists are endowed with stocks of physical capital and the homogeneous human capital as well. Capitalists produce good B with machines and workers in the firms they own. Technology is given and known to all. Two other factors that affect output are also considered as having a given value: length of the working period (say, 40 hours per week) and degree of effort intensity put in by workers.

Therefore, capitalists know that the quantity of output per unit of time will depend upon the quantity of workers hired; they also know that more workers will produce more output, but double number of workers will not produce double quantity of output, but less. This is due to the assumption that production is subject to *diminishing returns* of labor inputs. Diminishing returns implies that average productivity of labor declines as more workers are hired. The additional output that results from additional worker is called *marginal productivity of labor.* Diminishing returns also implies that the marginal productivity of labor declines as more workers are hired. Output per worker can also be called *average productivity of labor*. An important principle is the following: if marginal productivity of labor declines, then average productivity of labor will be above the marginal and will also decline.

Firms seek to maximize total profits. Profits are equal to the difference between total revenues and total costs. Therefore, profit maximization implies maximizing total revenues and minimizing total costs. Total revenues are equal to the price of good B multiplied by the quantities of good B sold; total costs are equal to the wage bill (nominal wage rate multiplied by the quantity of workers hired) plus the cost of depreciation of physical capital.

The cost of depreciation is a fixed one, for the model assumes that it is a fixed proportion of the value of the capital stock of the firm. The cost of depreciation allows the firm to maintain the stock of capital constant period after period. Depreciation is what makes production a process. Therefore, in what follows profits will be net profits, net of depreciation, whereas total output will also be net output, net of depreciation.

Consider now the behavior of capitalists taken individually. Individual capitalists are price takers in both markets: good B and labor. Because they operate in Walrasian markets, they can sell all the quantities of good B and buy all the quantities of labor that they are willing to exchange at the given market prices. Capitalists do not hold cash balances. There is no need for working capital, for they pay wages at the end of the production period.

Each firm will seek to hire the number of workers that makes profits the largest, that is, the number that makes the difference between revenues and costs the largest. The behavior of the individual capitalist j can be represented by the following structural equations:

$$\text{Maximize } P_j' = P_b\, Q_{bj} - P_h\, D_{hj} \tag{3.3}$$

subject to the production function

$$Q_{bj} = \Phi_j\,(D_{hj}, K_{bj}),\ \Phi_{j1}>0,\ \Phi_{j11}<0$$

The *production function* shows the input-output relation in the production process. Although technological knowledge is a public good, firms differ in their production functions on two accounts: differences in the entrepreneurial talents of capitalists and differences in the firm's location and access to public infrastructure. System (3.3) is a mathematical problem that the capitalist will seek to solve by different methods, such as intuition, trial and error, and so on. The assumption is that, whatever the method used, the solution found by the capitalist will be equal to the mathematical solution. Therefore, the capitalist behaves *as if* he solved this mathematical problem.

The mathematical solution indicates the equilibrium condition, which is the following:

$$\Phi_{j1}\,(D_{hj}, K_{bj})\,P_b = P_h \text{ or} \qquad (3.4)$$
$$\Phi_{j1}\,(D_{hj}, K_{bj}) = w.$$

The first equation indicates that profit is maximized when the value of the physical marginal productivity of labor is equal to the nominal wage rate. The second is obtained just by dividing the first by the price level (P_b), and indicates that profit is also maximized when the physical marginal productivity of labor is equal to the real wage rate. Either equation has only one variable to solve for: the quantity of labor demanded by the firm (D_{hj}), which is the number of workers that will make profits the largest.

According to the model, labor productivity is an important factor in the decision of firms to hire workers. It should be emphasized, however, that productivity is not a personal characteristic of the worker, something that could be written in his or her ID. The same worker accompanied by 30 other workers will result in a certain quantity of total output, which will be different if accompanied by 60 workers. Labor productivity is a collective result. However, labor productivity will also vary with different quantities of machines, or with different technology incorporated in the machines. In short, given the level of technology, labor productivity depends upon the number of workers and the stock of capital, as shown in equation (3.4).

The equilibrium condition is stable. If for some reason the firm has hired more workers than the optimal number, the nominal wage rate will be higher than the value of the physical marginal productivity of labor; hence, profits are not being maximized; then firms will reduce workers to increase the value of the physical marginal productivity (by the diminishing returns assumption) and the equilibrium number of workers will be restored. If for some reason the firm has hired fewer workers than the optimal number, by the same argument, firms will seek to hire more workers and the equilibrium

number of workers will also be restored. Therefore, the comparative statics method can be used to determine the effects of changes in the values of the exogenous variables upon the endogenous variable: the quantity of labor demanded.

If the price level goes up, then the value of the marginal productivity of labor will increase and it will become higher than the nominal wage rate; thus, the quantity of labor hired by the firm will increase. If the nominal wage rate rises, the value of the marginal productivity of labor will become smaller than the nominal wage, then the quantity of workers hired by the firm will decrease. Finally, if the capital stock of the firm increases, then the value of the marginal productivity of labor will rise (because workers will be equipped with more machines) and will become higher than the nominal wage rate; hence, the quantity of labor hired will increase. The individual firm's labor demand function can then be written as follows:

$$D_{bj} = h^j (P_b, P_b, K_{bj}) = H^j (w, K_{bj}),\ H^j{}_1<0,\ H^j{}_2>0 \tag{3.5}$$

subject to the firm's budget constraint $P_b S_{bj} = P_b Q_{bj} - P'_j = P_b D_{bj.}$

The constraint just says that the firm's value of market supply of good B (equal to total revenue minus profits) must be equal to the value of the payroll.

Market General Equilibrium

Workers and capitalists interact through market exchange. The result of such interactions is the object of general equilibrium analysis. There are three markets in this model: labor, good B, and money. Workers are willing to sell labor and buy good B; capitalists are willing to buy labor and sell good B. In addition, workers need to hold cash balances, and the supply of money comes exogenously from the government.

The individual behavior of a typical worker was presented earlier. The aggregate demand function for good B and the aggregate demand function for money are then obtained just by adding the individual demand functions. There are no direct interactions between workers, such as collective action; interactions among workers are only indirect, through markets. Hence, by aggregating equations (3.1) and (3.2) over all workers, we get

$$D_b = F (w, S_b),\ F_1>0,\ F_2>0 \tag{3.6}$$
$$D_m = P_b L (D_b),\ L'>0,\ L''<0$$

subject to the aggregate budget constraint $P_b S_b = P_b D_b + (D_m - S_m)$.

The individual behavior of firms was also presented earlier. There are no direct interactions between capitalists either, such as collective action; interactions among capitalists are only indirect, through markets. Hence, the aggregate demand function for labor is obtained just by adding the individual function, shown as equation (3.5), over all firms, as follows:

$$D_h = H(w, K_b),\ H_1<0,\ H_2>0 \tag{3.7}$$

subject to the aggregate budget constraint $P_b S_b = P_h D_h$.

In the marketplace, workers and capitalists are price takers. Prices are exogenous to them. No one has the power to set the market prices. How are then market prices determined? Prices are determined by the interaction of all social actors. The equilibrium prices will be those that clear the markets: people willing to exchange quantities at those prices will be able to do so; if not, prices will increase or decrease, depending on excess demand or excess supply situations. Therefore, prices and quantities in the market system are endogenous and their equilibrium values are determined by the interactions between buyers and sellers.

The general equilibrium conditions require that excess demand be equal to zero in each particular market. Then

$$\begin{aligned} &\text{Labor market } S_h = D_h \\ &\text{Commodity market } S_b = D_b \\ &\text{Money market } S_m = \mathrm{D}_m \end{aligned} \tag{3.8}$$

subject to the aggregate budget constraint

$$P_b (D_b - S_b) + P_h (D_h - S_h) + (D_m - S_m) = 0.$$

The aggregate budget constraint, which is also a general equilibrium condition, is obtained just by adding up the budget constraints of the two social groups, workers and capitalists.

How does the competitive market system solve for prices and quantities? There are three markets and three nominal prices and three quantities to solve for, a total of six endogenous variables. There are two conditions in each market, a demand function and a supply function, a total of six conditions. The system thus contains equal number of equations and unknowns.

In this particular neoclassical model, the unknowns are less than six. For one thing, the nominal price of money is one (the price of a dollar in dollars

is one); two quantities are exogenously determined: quantity of employment and quantity of money. In addition, from the aggregate budget constraint, it follows that if two of the three markets are in equilibrium, the third will necessarily be in equilibrium too. This is known as the *Walras Law of Markets*, which is a theorem derived from the assumption that all markets are Walrasian.

Because the three markets are Walrasian, one of them can be eliminated from the general equilibrium analysis. We can therefore choose to eliminate the good B market as the redundant one; thus, it is sufficient to solve the labor and money markets to reach general equilibrium. Because the quantities in these markets are exogenously determined, the system now has only two unknowns (the price level and the nominal wage rate) and two conditions (the demand function for labor and the demand function for money). Therefore, there are equal number of conditions to be satisfied and variables to be solved, that is, equal number of equations and unknowns in mathematical terms, which implies that in principle the solution should exist.

The assumption is that the market system is able to solve this problem. This is the general market theory. The market system works *as if* it solved a system of equations. In the real world the market system may use particular mechanisms to solve for prices and quantities (e.g., auctioneers, by costless trial and error), but these mechanisms are supposed to be equivalent to solving a system of equations.

The exogenous and endogenous variables of the market system are:

Exogenous variables: S_b, S_m, K_b
Endogenous variables: P_b, P_h, w, $D_b = S_b = Y$, P, W.

In this particular model of the neoclassical theory, the general equilibrium solution is very simple. It can be shown graphically.

Figure 3.1 depicts the labor market in panel (a) and the money market in panel (b). In the labor market, due to the assumption of diminishing returns, physical marginal labor productivity decreases as more workers are employed to work with a fixed stock of capital (K), as shown by curve H, which also represents the labor demand curve. Given the quantity of labor supplied, the labor market determines the real wage rate. Equilibrium is with full employment. Once the real wage rate is known, the real wage bill will also be known, which will determine the level of the money demand curve, which in turn will determine the price level in the money market. Finally, once the price level is known, the nominal wage rate will also be determined so as to be consistent with the real wage rate of equilibrium.

Notice that the money market solution is determined once the labor market solution has been determined. The market system comprises two subsystems: the real one (the labor market) and the nominal one (the money market). In this particular model, the labor market constitutes the *core of the general equilibrium*: it is sufficient to solve this market because the rest of endogenous variables can be resolved just by implication.

Production and distribution of equilibrium can then be represented by the following equations:

$$Y = P + W = P + w\, S_h \tag{3.9}$$

Total net output or national income of equilibrium (Y) is determined by quantity of employment, which in equilibrium is equal to the exogenous quantity of labor supply. Total income is distributed to capitalists as profits (P) and to workers as wages (W), which in turn is equal to the real wage rate multiplied by the quantity of employment. Income distribution of equilibrium can be seen in figure 3.1 panel (a), where total income is equal to the area under the marginal labor productive curve, the curve HE, which is distributed to wage bill and profits.

According to the neoclassical model, this is how the market system operates. The model says that whatever the market clearing mechanism used by the markets in the real world is, it is equivalent to (not equal to) what the model says. The market solution is the equilibrium situation in the sense that nobody has the power and the will to change the solution. It is static equilibrium in the sense that as long as the values of the exogenous variables remain fixed, the equilibrium values of the endogenous variables will be repeated period after period.

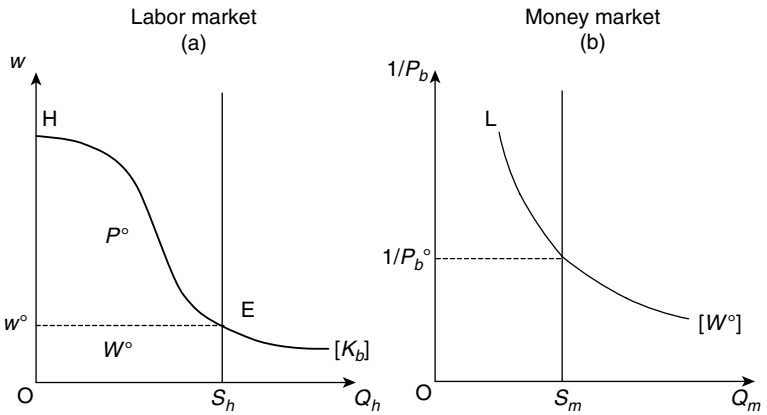


Figure 3.1 Neoclassical general equilibrium.

Empirical Predictions

The general equilibrium is stable because the labor market (the core of general equilibrium) is stable. Therefore, beta propositions can be determined by using comparative statics. There are three exogenous variables in the market system: quantity of labor supplied, capital stock, and quantity of money supply. Now it is time to show the effect of changes in these exogenous variables upon the equilibrium values of the endogenous variables. Because the interest is in the short-run equilibrium, the only effect to be analyzed is changes in the money supply, maintaining fixed the stocks of capital and labor.

Suppose the government increases the money supply (say by helicopter or any similar mechanism), this increase will directly generate an excess of real cash balances willingly held by workers. They will then seek to get rid of the excess by buying a higher quantity of good B, which will tend to raise the price level. A higher price level would induce firms to produce more and hire more labor. Given that workers are fully employed, the competition of firms for workers would lead to a higher nominal wage rate, which would keep rising until the real wage and the employment of equilibrium are restored at the initial equilibrium. Total output and real wage will then remain constant, and so will income distribution.

The price level will rise in proportion to the increase in the money supply, which will reduce the real cash balance to its original equilibrium level. The real cash balance will not change. Hence, the new equilibrium will have the real variables unchanged; the only effects of an increase in money supply are on the nominal variables: both the price level and the nominal wage rate will increase in the same proportion that money supply did. If the money supply doubles, both nominal prices will double, leaving both the real wage rate and the real cash balance unchanged. There is no effect of money supply on real variables; then *money is neutral* in this model.

A final comment on the aggregation problem is in order. If money stock increases for the *individual worker alone*, as we mentioned earlier, he will seek to eliminate this excess by buying extra quantities of good B in the adjustment period only, which implies a negligible effect over several periods. Now it is found that if money stock increases for *everyone*, no one would be able to buy more goods even in the adjustment period. This is a clear case of *fallacy of composition*: what is true for the individual is not true for the aggregate. However, at the aggregate level, in their attempt to restore real cash balances, individuals end up raising the price level, and in the same proportion of the money increase; as a result, the aggregate quantity demanded of good B does not change. Indeed, we introduced above the assumption of ignoring the real cash balance in

the individual demand for good B just by anticipating this fallacy at the aggregate level.

The Keynesian Society

Another abstract society will now be presented. This will be called the Keynesian society. This construction follows the work of the English economist John Maynard Keynes (1936).

In this society, not all markets will be Walrasian. In some markets, nominal prices will be exogenously determined. Some buyers or sellers will then be unable to realize the market exchange in the quantities that they are willing to exchange at the market price; so the market will operate with excess demand or excess supply. This type of market is called a *non-Walrasian market*.

The primary assumptions of the Keynesian theory (K) can be presented as the following set of alpha propositions:

- $\alpha(K).(1)$ *Institutional context*: (a) Rules: People participating in the economic process are endowed with economic assets, which are subject to private property rights; they exchange goods under the norms of market exchange, which include the social norm that nominal wage rates cannot fall. Walrasian and non-Walrasian markets constitute the market system; the labor market is non-Walrasian. The political regime is democratic. (b) Organizations: firms, households, and the government.
- $\alpha(K).(2)$ *Initial conditions*: Factor endowments of society (capital per worker) are such that the productivity of total labor is high enough for firms to pay wages and generate profits. The exogenously given nominal wage rate is above the Walrasian nominal wage rate.
- $\alpha(K).(3)$ *Economic Rationality of Agents*: Individuals act motivated by self-interest.

A Keynesian Model: The Short Run

In order to make Keynesian theory falsifiable, a Keynesian model must be constructed, which can be seen as the task of constructing a particular stage set in which social actors will play their roles. The following auxiliary assumptions are introduced for that purpose:

- There are two social groups: capitalists who are endowed with stocks of physical capital and human capital, and workers with stocks of human capital and cash balances.

- Government behavior consists of supplying money.
- Three markets will constitute the market system: commodity and money are Walrasian, and labor is non-Walrasian. Workers are endowed with the same human capital. Hence there is only one labor market. Only one good is produced in society, called good B. Workers and capitalist firms exchange labor services and goods in the labor and commodity markets; money supplied by the government is used as the means of payment and is exogenously determined. Wages are paid at the end of the production period; consequently, workers need to hold cash balances for transaction purposes. There is no market for renting capital services.
- Capitalists and workers interact and compete in the market to obtain their objectives of self-interest, but none has the power to set prices (perfect competition); they all face costless information on market prices and technology. Labor productivity levels vary across firms.
- The economic process is static; it is closed as well, for market exchange with other societies is ignored.

The analysis of the behavior of social actors will start with that of workers. The rationality of the typical worker consists of seeking to maximize her real income, the quantity of good B, subject to her budget constraint. Workers know that the labor market functions with job rationing. Not all workers will be able to exchange the amount of labor they are willing to sell. Once firms have determined the total quantity of labor demanded (D_h), at the given nominal wage rate, the quantity of labor that the capitalists will buy from each household is exogenously determined. In the labor market, there is a random rationing mechanism, exogenously determined, such that firms will buy D_{hi} units of labor from household i, where $D_{hi} = 0$ for some households.

The individual worker (i) seeks to maximize real income subject to the budget constraints. Thus the set of structural equations is

$$P_h D_{hi} = P_b D_{bi} + (D_{mi} - S_{mi}) \qquad (3.10)$$
$$D_{mi} = P_b L_i (D_{bi}),\ L_i' > 0,\ L_i'' < 0.$$

In this model, $D_{hi} = n_i D_h$, where n_i is exogenously given and represents the share of worker i in total employment D_h, which is the total quantity demanded by capitalists at the given nominal wage rate P_h. When $n_i = 0$, the worker i is unemployed. The first budget constraint reflects the macrofoundations of microbehavior, that is, the assumption that the individual worker

operates in a non-Walrasian labor market. The equilibrium condition is that the total money supply is willingly held by workers ($S_{mi} - D_{mi} = 0$). This equilibrium is (trivially) stable and then comparative statics may be applied to derive beta propositions on the individual behavior of workers.

An increase in D_b will increase the worker's chances to get a job and thus its effect will be positive on the quantity of good B demanded. An increase in the real wage, due to changes in the nominal wage rate or the price level, will also be positive. Changes in the worker's labor supply will be ignored; changes in the aggregate labor supply will be introduced later on. A change in money endowment will increase the worker's real cash balances, which she will seek to reduce by buying extra quantities of good in the period of adjustment only. After this adjustment is made, she will consume the initial quantities of good B and hold the same real cash balances, period after period; therefore, over several periods, the real cash balance effect can be ignored.

The reduced form equations are thus the following:

$$D_{bi} = F^i (P_b, D_b, P_h), \quad F_1<0, F_2>0, F_3>0 \qquad (3.11)$$
$$D_{mi} = P_b L_i (D_{bi}), \qquad L_i'>0, L_i''<0.$$

The first equation shows the quantity of good B demanded by the individual worker and the second shows her demand for cash balances.

On the behavior of the typical capitalist, the demand for labor and the supply of good B functions will be equal to those derived for the neoclassical model. The reason is that firms face the same exogenous variables as in the neoclassical model. The only difference is that these functions cannot be expressed in terms of real wages, for the nominal wage rate is exogenous. Thus, for capitalist *j*, the reduced form equations are

$$D_{hj} = H (P_b, P_h, K_{bj}), H_1>0, H_2<0, H_3>0 \qquad (3.12)$$
subject to $P_b S_{bj} = P_h D_{hj.}$

The first equation shows the labor demand function of firm *j*, expressed in nominal prices. The second equation shows the budget constraint of the firm.

Market General Equilibrium

Aggregating the equations in (3.11) over all workers, the following market demand functions are derived:

$$D_b = F(P_b, D_b, P_h),\ F_1<0,\ F_2>0,\ F_3>0 \qquad (3.13)$$
$$D_m = P_b\, L(W) = M(P_b, D_b, P_h),\ M_i>0$$
$$\text{subject to } P_h\, D_h = P_b\, D_b + (D_m - S_m).$$

The first equation shows that the aggregate quantity of good B demanded depends on nominal prices and on the employment level. This is the result of assuming a non-Walrasian labor market, in which the income of workers does not depend on the quantity of labor they are willing to sell, but on the quantity of labor the firms are willing to buy. Although, the demand function F is homogeneous of degree zero in nominal prices, it cannot be written in terms of real wages. When the market nominal wage rate is fixed, there is no economic meaning in the statement "if both nominal prices double, the quantity demanded will remain unchanged."

The second equation shows that the quantity of cash balance demanded depends on the price level and on the real wage bill because, in the aggregate, the quantity of good B demanded is identical to the real wage bill (W). This demand function can, in turn, be expressed as depending upon the price level, the employment level, and the nominal wage rate. The effects of the employment level and the nominal wage rate are both positive.

The effect of the price level is more involved. A higher price level has two effects that run in opposite directions. One effect is positive because at a higher price level more cash balances will be needed; but a higher price level reduces the real wage bill and thus less cash balances will be needed. The first effect dominates over the second because it is direct and proportional, whereas the second effect is smaller due to the assumption that income elasticity is less than one; thus the net effect of the price level on the demand for nominal cash balance is positive.

Aggregating equation (3.12) over all firms, the market demand function for labor will become

$$D_h = H(P_b; P_h, K_b),\ H_1>0,\ H_2<0,\ H_3>0 \qquad (3.14)$$
$$\text{subject to } P_b\, S_b = P_h D_h.$$

The quantity of labor demanded depends positively on the price level, for a higher price level implies a higher value of the marginal productivity of labor, and negatively on the nominal wage rate, for a higher nominal wage implies the need to increase the value of the marginal productivity of labor, which implies less employment (diminishing returns dictates this result); and positively on the stock of physical capital, for a higher stock implies a higher marginal productivity of labor. The constraint says that what firms

pay as wages must be equal to the value of what they supply to the good B market. Since workers spend what they earn (savings are nil), savings can come only from profits.

From these structural equations, the general equilibrium conditions of the market system can be written as

$$\text{Labor market } S_h \equiv D_h + U,\ P_h \geq P_h^* \tag{3.15}$$
$$\text{Good B market } S_b = D_b$$
$$\text{Money market } S_m = D_m$$

subject to the aggregate budget constraint

$$P_h D_h + P_b S_b = P_b D_b + P_h D_h + (D_m - S_m) \text{ or}$$
$$P_b (D_b - S_b) + (D_m - S_m) = 0.$$

Unemployment (U) is now a possible outcome of the economic process. General equilibrium includes the condition that the nominal wage rate cannot be smaller than P_h^*, which is the current nominal wage, exogenously determined. Nominal wages are sticky downward.

Given the values of the exogenous variables, the interaction between capitalists and workers in the marketplace will determine the equilibrium values of the endogenous variables of the system, which will be repeated period after period. These variables are

Exogenous variables: $S_m,\ P_h^*,\ K_b,\ S_h$

Endogenous variables: $w,\ D_h,\ U,\ S_b = D_b = Y,\ P,\ W,\ P_b,\ P_h.$

It should be noted that variables that were exogenous in the microeconomic equilibrium are endogenous in general equilibrium, such as the employment level and the price level. The market mechanism will be able to solve the general equilibrium problem. The theory of markets is also applied to this Keynesian model: The market system operates *as if* it solved a system of equations, but it is a system of equations that is different from that of the neoclassical model.

The core of the general equilibrium must be determined now. No subsystems exist here; hence, the solution is simultaneous. In non-Walrasian markets, the values of the quantities supplied and demanded are identical, and they cancel out in the aggregate budget constraint. Walras's law applies to Walrasian markets only. So, of the two Walrasian markets, one is redundant. Let the commodity market be redundant. Thus, both the labor and

money markets constitute the core of the system, which can solve for the two endogenous variables P_b and D_h. The core of the system includes

$$D_h = H(P_b; P_h, K_b),\ H_1>0,\ H_2<0,\ H_3>0 \qquad (3.16)$$
$$S_m = M(P_b, D_h; P_h),\ M_i>0.$$

The first equation shows the labor demand function, whereas the second shows the equilibrium condition in the money market. The solution is simple enough to be shown graphically.

Figure 3.2 displays the core of the general equilibrium. Panel (a) shows the labor market, whereas panel (b) the money market. In order to solve for

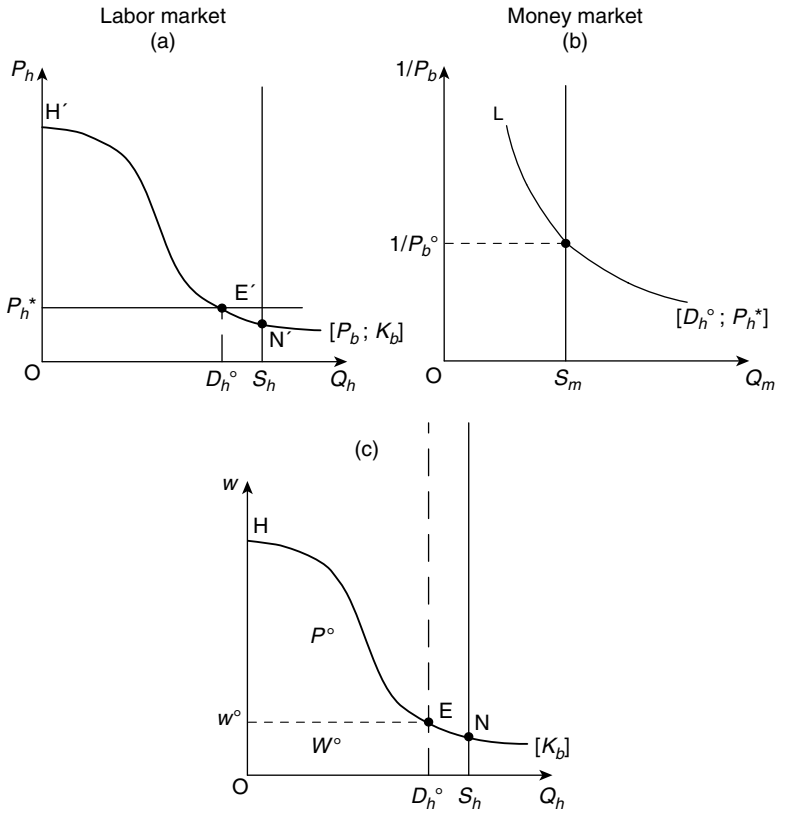


Figure 3.2 Keynesian general equilibrium.

the price level, the employment level must be determined; conversely, in order to solve for the employment level, the price level must be known. The solution is simultaneous.

The other endogenous variables are determined from the core solution just by implication. Thus, the core of the general equilibrium implies an equilibrium real wage rate (w°), which is shown in panel (c). The market real wage rate is higher than the Walrasian wage rate (indicated by the point N). General equilibrium with unemployment is the outcome. Total output and its distribution between workers and capitalists are also determined, as can be seen in panel (c). It can also be shown there that in fact the commodity market is in equilibrium. The quantity of good demanded is equal to the wage bill; the quantity supplied is equal to total output (the area under the segment HE) minus profits, which is represented by the same area of the wage bill.

The equilibrium net output or national income (Y) and its distribution may be represented as follows:

$$Y = W + P = w D_h + P \tag{3.17}$$

Once the general equilibrium is reached, the equilibrium values of the endogenous variables will be repeated period after period as long as the exogenous variables remain fixed. This is an *equilibrium* situation, even with the existence of unemployment, because no agent has both the power and the incentive to change the situation.

Empirical Predictions

The core of the static general equilibrium is stable. The reason is that in the labor market (a non-Walrasian market) stability is trivial, as firms will be able to readjust automatically if the quantity of employment is for some reason out of equilibrium, that is, there cannot exist instability in the labor market. Therefore, the only place where instability would possibly exist is the money market (a Walrasian market). The money market is also stable here because the demand curve is sloping downward and the supply curve is a vertical line; hence, any price level that is for some reason out of equilibrium would restore its equilibrium automatically. In this very simple model, stability of the general equilibrium requires stability in the money market only, which is fulfilled. Therefore, comparative statics can be applied to the general equilibrium solution to derive beta propositions.

The short-run model, in which the productive capacity of society is given, will only show the effects of changes in money supply and nominal

wage rate upon the main endogenous variables of the core: employment level and price level.

If the stock of money increases, the direct effect will be to increase both the price level and the employment level. The reason for this is that workers will try to get rid of the excess cash balances by buying more quantities of output, which will tend to increase the price level. Firms would then have incentives to produce more and hire more workers, which is viable because there is unemployment. Nevertheless, that will not be all; some additional effects may take place. The real wage rate would fall, which would then tend to change the real wage bill, which in turn would change the demand for money. However, the change in the real wage bill is ambiguous and would have a small effect or none upon the demand for money and on the price level; then the initial direct effect will prevail.

This effect can be visualized in figure 3.2 panel (c). An increase in money supply raises the price level, which in turn reduces the real wage rate, which implies a rise in the employment level. So the initial equilibrium situation, at point E, will move to another equilibrium situation, to a point that lies below point E, along the labor demand curve HN. Unemployment is thus reduced. Total output increases. The change in income inequality is ambiguous because the change in the real wage bill is too.

This Keynesian model also predicts that a sufficiently large increase in the supply of money will ultimately result in full employment in the labor market. Equilibrium could move from point E to point N in figure 3.2 panel (c). Full employment can be attained through government policies. If money supply is increased further, the initial effect will be to increase the price level, which induces firms to hire more labor; but there are no unemployed workers to hire; hence, firms will compete for workers and the nominal wage rate will rise. The ultimate effect of an increase in money supply will be to increase the price level and the nominal wage rate in the same proportion, leaving unchanged the real wage rate. Under full employment, money becomes neutral, as in the neoclassical model.

The effect of an exogenous increase in the nominal wage rate will directly increase the real wage rate, which will lead firms to reduce the quantity demanded for labor. Firms will tend to reduce employment and thus total output will tend to fall. The wage bill change is ambiguous; therefore, the demand for money and the additional effect upon the price level will also be ambiguous; hence, the direct effect will prevail. In figure 3.2 panel (c), the initial equilibrium situation at point E will move to another equilibrium situation, say to a point E″that lies above point E, along the labor demand curve. Unemployment increases. Total output falls. Change in income inequality is ambiguous.

The labor market in this Keynesian model works as follows. Whenever there is an excess labor demand situation, the nominal wage will tend to rise (as in a Walrasian market); but whenever there is an excess labor supply situation, the nominal wage cannot fall, and the situation will remain unchanged as the equilibrium situation. Nominal wages are sticky downward, but not upward.

The Keynesian model presented here is a very simple one. Macroeconomic textbooks present Keynesian models that are more complex (Blanchard 2009). For example, introducing the bond market, the government can increase money by buying bonds from the public at a price that clears the market; moreover, the bond price and the interest rate are inversely related, that is, if the bond price rises, the interest rate falls. This way of increasing money is called "open market operation," while the mechanism utilized in the model presented earlier is known as "helicopter money," for money is injected directly into the economy as if by using a helicopter. Therefore, in the textbook model, the government can choose as policy instrument either money supply or nominal interest rate, that is, there is only one degree of freedom: once one of them is chosen, the other will be determined endogenously. The essential relations between nominal and real variables are thus captured in the simpler model presented here.

The Classical Society

The classical society will also be presented as an abstract society. The works of David Ricardo (1821) and Karl Marx (1867) inspire this construction.

Property structure is such that a classical society is a class society. There are two social classes: capitalists who are endowed with physical capital and workers who are endowed with human capital alone. Capitalists do not rent out the services of their capital in the market because they wish to avoid the formation of new firms and new capitalists. Firms can only be created by accumulating physical capital.

In the classical society, the labor market operates with the real wage rate of subsistence (w^*), which satisfies the worker's subsistence needs, which is exogenously given. The theory assumes that at the subsistence real wage rate, there is excess labor supply. The labor market is thus non-Walrasian; the price of labor services cannot change to clear the labor market.

Because capitalists control the ownership of physical capital, capitalists are able to set the length of the working time at a level that is above of what would be needed to produce the quantity of subsistence goods for workers. This length is a parameter of the production function. Thus, the total work length utilized by the firm (call it η) has two components: the necessary labor (η^*), which is that part of the length of the working time that

would be enough to produce the quantity of workers' subsistence goods, and the surplus labor (η^e), which is that part that produces surplus output. Capitalists appropriate in the form of profits the output produced with the surplus labor. This appropriation is called *exploitation*. Profits come from exploitation. Without surplus labor, profits could not exist.

To illustrate these concepts, suppose a society produces 100 kg of output per worker in a week, with 40 hours as the length of working time; also suppose the subsistence income per worker is equal to 50 kg per week. The production process can be conceptually separated into "necessary labor," which is equal to 20 hours (the time necessary to produce 50 kg for the subsistence of the worker), and the "surplus labor," the other 20 hours, in which the other 50 kgs will be produced. The necessary labor gives rise to wages, whereas the surplus labor generates profits.

This example shows that labor productivity must be sufficiently high to generate output per worker beyond subsistence wage. If labor productivity were 50 kg, this society could not generate profits and could not operate as a capitalist society. If labor productivity were less than 50 kg, this society could not be economically viable, not only under capitalism, but under any type of social organization; it could not exist.

The primary assumptions of the classical theory (CL) can be expressed as the following set of alpha propositions:

α*(CL). (1) Institutional context*: (a) Rules: People participating in the economic process are endowed with economic assets, which are subject to private property rights; people exchange goods subject to the norms of market exchange, which includes the norm that market real wages cannot be smaller than the cost of the worker's subsistence basket of goods. The market system operates with Walrasian and non-Walrasian markets; the labor market is non-Walrasian. (b) Organizations: Firms, households, and the government.

α*(CL). (2) Initial conditions*: Individuals are endowed with unequal quantities of economic assets. A social group concentrates the total stock of physical capital; there are two social classes: workers and capitalists.[1] At the initial real wage rate, which is exogenously determined, there is excess labor supply.

α*(CL). (3) Economic rationality of agents*: Individuals act motivated by self-interest. Capitalists seek two particular objectives: profit maximization and the preservation of their class position. Capitalists do not rent out the services of their capital stocks in the market for they prefer profits to rents. Capitalists seek to set the length of the working day above the socially necessary working time.

A Classical Model: The Short Run

In order to generate beta propositions, a classical model is constructed with the introduction of consistent auxiliary assumptions. They are:

- There are two social groups: capitalists who are endowed with stocks of physical capital and human capital, and workers with stocks of human capital and cash balances. Government behavior consists of supplying money to the economic process.
- Wages are paid at the end of the production period; consequently, workers need to hold cash balances for transaction purposes.
- Three markets will constitute the market system: commodity and money, which are Walrasian, and labor, which is non-Walrasian. Human capital is the same for all, so there is only one labor market. One sole good, called good B, is produced in society. Workers and capitalist firms exchange labor services and goods in the labor and the commodity markets; money supplied by the government is used as the means of payment and is exogenously determined. Capitalists and workers interact and compete in the market to obtain their objectives of self-interest, but none has the power to set prices (a perfect competitive market); they all face costless information on market prices and technology. Labor productivity levels vary across firms.
- The economic process is static and closed.

The study of the general equilibrium starts with the behavior of workers. Workers exchange their labor power in a labor market in which not all workers can get the amount of jobs they are willing to exchange at the prevailing real wage rate. In the labor market, there is a rationing mechanism, which is assumed to be random. Given the real wage rate, the quantity of labor demanded is known (D_b), and they will hire D_{bi} units of labor from each household i, where $D_{bi} = 0$ for some households.

In this context, the typical worker i seeks to maximize his real income, subject to the following constraints:

$$w\, D_{bi} = D_{bi} + (D_{mi} / P_b - S_{mi} / P_b) \qquad (3.18)$$
$$D_{mi} / P_b = L_i\,(D_{bi}),\ L_i' > 0,\ L_i'' < 0.$$

Here $D_{bi} = n_i D_b$, where n_i is the share of worker i in total labor demand D_b. When $n_i = 0$, the worker is unemployed. The first equation of the budget constraint reflects the macrofoundations of microbehavior: the worker operates in a context of non-Walrasian labor market and Walrasian commodity

and money markets. The real income of the individual worker depends on the quantity of labor bought by firms, not on the quantity sold by the worker (as is the case in the Walrasian labor market). The second equation shows the real cash balance constraint.

From the structural equations shown above, it follows that the equilibrium condition is that the supply of money must be willingly held by workers ($D_{mi} - S_{mi} = 0$). This condition is (trivially) stable. Comparative statics may now be applied to derive beta propositions: the effect of the exogenous variables (w, $D_{h,}$ P_h, P_b, S_{mi}) upon the endogenous variables (D_{bi}, D_{mi}).

An increase in the employment level will make more likely the employment of the individual worker. An increase in the stock of money endowments will lead the worker to get rid of the extra money by buying extra quantities of goods to restore his equilibrium real cash balance. As in the neoclassical model, it is assumed that the real cash balance effect at the individual level is negligible and can safely be ignored. Changes in labor supply are relevant at the aggregate level only. The effects of the other exogenous variables are similar to those obtained in the neoclassical model. The individual demand functions for good B and for money can then be written as the following equations:

$$D_{bi} = F^i(w, D_h),\ F^i_1 > 0,\ F^i_2 > 0 \qquad (3.19)$$
$$D_{mi} = P_b\, L_i\,(D_{bi}),\ L_i' > 0,\ L_i'' < 0.$$

On the other hand, the individual firm j seeks to maximize profits, subject to its individual endowments of capital and production function. For firm j we can write

$$\text{Maximize } P_j' = P_b\, Q_{bj} - P_h\, D_{hj} \qquad (3.20)$$
$$\text{subject to } Q_{bj} = \eta\, q_{bj} = \eta\, \varphi_j\,(D_{hj}, K_{bj}).$$

The production function of the firm refers to output q_b, which represents the output in the unitary period of production, say the hour. This unitary period can be replicated for η periods (say, 40 hours per week) and the firm will be able to produce proportionally more output, that is, the production function is homogeneous of degree one with respect to time. The work length or duration of the economic process η is exogenously given and is subject to $\eta \leq \eta'$, where η' is the maximum duration that is physically viable for workers (say, 60 hours per week). Hence, Q_{bj} is the flow of output per week.

From the structural equations shown above, it follows that the firm's equilibrium condition is that the marginal productivity of labor should be equal to the real wage rate, that is

$$\eta\ \varphi_{j1}\ (D_{hj}, K_{bj}) = w \tag{3.21}$$

Because the wage rate must be equal to the subsistence wage w^*, which is equal to the output generated with the necessary labor, the following relations hold true for the typical firm j:

$$w = w^* = \eta^*\ (q/D_h)_j = \eta\ \varphi_{j1}\ (D_{hj}, K_{bj}) < \eta\ (q/D_h)_j \tag{3.22}$$

The first equality indicates that the market wage rate must be equal to the subsistence wage; the second equality that, given the hourly average productivity of labor and given the subsistence wage rate, the necessary labor η^* becomes determined endogenously (say 20 hours per week); the third equality shows just the profit maximization condition, stated above; and the last inequality shows that the marginal productivity of labor is smaller than the average productivity of labor, which follows from the assumption of diminishing returns.

The set of relations shown in the system (3.22) implies that $\eta > \eta^*$ (say η is equal to 40 hours per week). Because $\eta=\eta^*+\eta^e$, this condition in turn implies that $\eta^e>0$. Hence, if there is no surplus labor, there will be no profits. This set of relations also says that positive profits imply that the weekly average productivity of labor must be higher than the weekly marginal productivity of labor. The labor exploitation account is therefore consistent with the labor productivity account in explaining the generation of profits.

The firm's equilibrium is (trivially) stable. Comparative statics can then be applied to derive beta propositions on the firm's behavior: the effect of changes in the values of the relevant exogenous variables (w, K_{bj}) upon the relevant endogenous variables (D_{hj}, P_j).

The quantity of labor demanded depends negatively upon the real wage rate and positively upon the stock of physical capital. The supply of good B must be equal to the real wage bill the firm must pay for the hired labor, as a condition of its budget constraint. Hence, the reduced form of the behavior of firm j can be written as

$$D_{hj} = H^j\ (w, K_{bj}),\ H_1<0,\ H_2>0 \tag{3.23}$$

subject to the budget constraint $S_{bj} = Q_{bj} - P_j = w\ D_{hj}$.

General Equilibrium

In order to construct a general equilibrium system, the structural equations showing market behavior of the social actors are needed. Aggregating over all workers, the system (3.18) becomes

$$D_b = F(w, D_h),\ F_1>0,\ F_2>0 \qquad (3.24)$$
$$D_m = P_b\, L(D_b),\ L'>0,\ L''<0$$
subject to the budget constraint: $w\, D_h = D_b + (D_m / P_b - S_m / P_b)$.

The particular feature of this model is that the quantity of good B demanded depends positively upon both the real wage rate and the quantity of labor demanded (not the quantity of labor supplied, as in the neoclassical model). The higher the employment level D_h, the higher the demand for good B.

Aggregating over all firms, equation (3.23) becomes

$$D_h = H(w, K_b),\ H_1<0,\ H_2>0 \qquad (3.25)$$
subject to the budget constraint: $S_b = w\, D_h$.

This is the aggregate labor demand function and the constraint shows the aggregate supply of good B, which is obtained directly from the aggregate budget constraint of firms.

The equilibrium conditions for general equilibrium are

$$\text{Labor market } S_h \equiv D_h + U,\ w \geq w^* \qquad (3.26)$$
$$\text{Commodity market } S_b = D_b$$
$$\text{Money market } S_m = D_m$$

subject to the aggregate budget constraint:

$$w\, D_h + S_b = D_b + (D_m / P_b - S_m / P_b) + w\, D_h \text{ or}$$
$$(D_b - S_b) + (D_m / P_b - S_m / P_b) = 0.$$

Unemployment (U) is now a possible outcome of the economic process because the market real wage cannot be smaller than the subsistence wage w^*.

Given the values of the exogenous variables, the interaction of workers and capitalists in the market system will determine the values of the endogenous variables of the system. These variables are

Exogenous variables: K_b, S_h, S_m, w^*, η

Endogenous variables: D_h, w, U, Q_b, P, W, P/W, P_b, P_h.

The assumption is that the market system will be able to solve this classical model, for it operates as if it solved a system of equations. However, the system to be solved is different from that of the neoclassical model explained earlier.

In order to determine the core of the model, take note that two Walrasian markets and one non-Walrasian market constitute the market system. Among Walrasian markets, one market is redundant, due to the Walrasian law; let good B market be the redundant one. In addition, two subsystems can be seen, a real (labor and commodity markets) and a nominal (money market); moreover, the monetary subsystem needs a real variable to reach a solution, but not vice-versa. Thus, the general equilibrium solution can be found by simply solving for the labor market and the money market, and solving them sequentially. Thus, the core of the general equilibrium system is composed of the labor market alone, which will solve for the employment level. The money market can then solve for the price level. The rest of the endogenous variables will be solved just by implications.

Figure 3.3 represents the classical general equilibrium. This classical model is sufficiently simple to show the solution graphically. The labor market is presented in panel (a). Curve HBC represent the average productivity of labor and curve HEN the marginal productivity of labor. The market real wage rate is equal to the subsistence wage rate (w^*), at which there

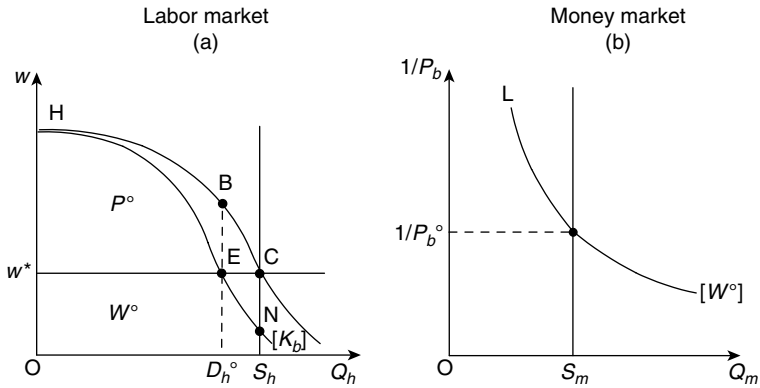


Figure 3.3 Classical general equilibrium.

exists excess labor supply. The employment level is determined by the labor demand curve HN. Point E shows the equilibrium situation in the labor market. (The Walrasian real wage rate is at point N, below the subsistence wage rate.) Total wage bill is then determined. Total profit is just equal to total output minus the wage bill, where total output is equal to the area under segment HE or, alternatively, the area determined by the value of B on the HBC curve multiplied by total employment.

Figure 3.3 also shows that the core of the general equilibrium can be represented in terms of labor value accounts. Thus

$$P/W = (Q - W)/W = (\eta\, q/D_h - \eta^* q/D_h)/\, \eta^* q/D_h \qquad (3.27)$$
$$= (\eta - \eta^*)/\eta^* = \eta^c/\eta^*.$$

The ratio of profits to wage bill is just equal to the ratio of surplus labor to necessary labor, which is called the *rate of exploitation*. In terms of labor productivity, the rate of exploitation is equal to the ratio average labor productivity to real wage rate (equal to the marginal productivity) minus one, that is, equal to the ratio $BE/D_h{}^0E$. Surplus labor underlies observed profits.[2]

The money market is represented in panel (b). Once the wage bill is known, the demand curve for money is determined, the curve L. Given the quantity of money supplied, the price level is determined by the money demand curve. So, equilibrium in the money market is determined sequentially once the equilibrium in the labor market has been solved. Changes in the labor market will affect the money market, but not vice versa.

The general equilibrium solution is with unemployment. Can this be an equilibrium situation? Classical theory intended to show the shortcomings of the capitalist system. This is a class conflict society, where workers if hired are exploited, but if not exploited are unemployed and have zero wage income. But the unemployment situation will generate social tensions. Indeed, Marx predicted the breakdown of the capitalist system because of this and other social tensions created in the economic process, especially during the calvaries of capitalism observed in its initial periods. However, reality has refuted this prediction.

A reason for the refutation may be found in the introduction of institutional innovations in most of the First World countries, such as unemployment insurance, which transfers incomes to the unemployed through the public budget. The amount of this transfer should be lower than the market wage rate so as to maintain the incentives for both employment seeking and labor discipline. Introducing the assumption that unemployment insurance exists, which provides the unemployed with an allowance that is lower than

the market wage rate, general equilibrium with unemployment is obtained. This situation can be repeated period after period because no social agent has the power and the will to change it.

From the solution of the core, other endogenous variables can be solved by implication. It can easily be verified in figure 3.3, panel (a), that the commodity market is also in equilibrium: the area showing the quantity demanded (the wage bill) is the same area showing the quantity supplied (total output minus profits). Once the price level is known, the nominal wage is endogenously determined, for it must be consistent with the real wage rate.

The net output that comes out of the production process is called the *economic surplus* (ES). According to the classical model presented here, workers get only subsistence wage in the equilibrium situation. Hence, capitalists appropriate the entire economic surplus in the form of profits. Thus,

$$ES = P = r\,K_b \tag{3.28}$$

Note that r is not a market price, but the average rate of return of capital.

Empirical Predictions

In this classical model, the general equilibrium is stable. Because the solution is sequential, it is sufficient to show that there exists stability in the equilibrium of the labor market, the core of the system, and in the money market, taken separately. That this is indeed the case can be seen in figure 3.3. The comparative statics method can then be applied to derive beta propositions. For the short-run analysis, the relevant exogenous variables are reduced to the money supply.

An increase in the aggregate money supply will affect the equilibrium real cash balances in the money market. Workers would seek to get rid of the excess cash balances by buying extra quantities of good B in the period of adjustment. Then the price level would tend to increase and thus firms would tend to produce more output and hire more labor, which could be materialized because there is unemployment. However, the subsistence real wage would tend to diminish, which by assumption is socially unviable. Then the nominal wage rate would have to increase and do so until the subsistence real wage is restored.

In the new equilibrium, all real variables in the labor market and commodity market will then remain unchanged. The only effect of a change in the money supply is to change the nominal variables; it does not affect the real variables; hence, *money is neutral.* However, it is so under equilibrium

with unemployment, a difference from the neoclassical and Keynesian models presented here.

In the classical society, the strict problem of distribution refers only to the distribution of surplus. In this particular model, the entire surplus goes to profits. More generally, the theory allows for models in which the economic surplus can be appropriated partly by capitalists as profits and partly by workers as higher-than-subsistence wage. Workers may appropriate part of the economic surplus depending on their relative strength in the distribution conflict. Thus, under classical theory not all the observed wage bill is part of the economic surplus; it includes only that part that exceeds the subsistence wage. Subsistence wage is the cost of maintaining workers, similar to the cost of maintaining horses in farms. It could not be part of the net output.

Finally, this classical model shows that a continuous exogenous increase in the stock of physical capital will also shift continuously the labor demand curve and will eventually eliminate the excess labor supply. Then there will eventually be excess demand for labor and the competition among firms for workers will imply equilibrium wage rates that are higher than the subsistence wage; hence, the real wage rate will endogenously rise. The same effect will result if the quantity of labor supplied decreases continuously. Under sufficient large changes in these exogenous variables, therefore, the labor market will become a Walrasian market, which means that this classical model may become the neoclassical model studied earlier.

Empirical Consistency: The First World Countries

From the list of eight empirical regularities about capitalism shown in chapter 2, two refer to the First World behavior in the short run: The existence and persistence of unemployment (Fact 1) and the interaction between nominal and real variables (Fact 4). Do the three theoretical models presented here explain these facts?

The empirical predictions of the neoclassical general equilibrium model state that the labor market equilibrium operates with full employment and that money is neutral, which Facts 1 and 4 refute. Although the neoclassical model is a system that is logically correct, it is empirically false.

The predictions of the Keynesian general equilibrium model are consistent with Fact 4 but refuted by Fact 1. With respect to Fact 1, although this model can predict the existence of equilibrium with unemployment, there are two logical problems. First, there is a logical problem with the existence of unemployment. The nominal wage rate is not truly exogenous, for it must

be set above the Walrasian wage rate if it is to generate unemployment initially. Second, the model also predicts equilibrium with full employment in the short run (which has never been achieved). The bottom line is that the existence of full employment would not refute the Keynesian model, that is, in the Keynesian model, unemployment is not a necessity; unemployment plays no role in the functioning of capitalism. In sum, the Keynesian model has difficulties in explaining Fact 1.

Regarding the classical theory, the general equilibrium model predicts that money is neutral even in the short run, which is refuted by Fact 4. As to Fact 1, the classical model faces the same type of logical difficulties shown for the Keynesian model. First, the existence of unemployment is exogenously determined. Second, the model also predicts full employment equilibrium in the long run, that is, the existence of full employment would not refute the model. Other classical models introduce mechanisms to maintain the "industrial reserve army" (unemployment) in the process of capital accumulation, so as to *avoid the increase in the real wage rate in the long run*, by assuming endogenous technological change that is labor saving, which displaces labor continuously. However, Fact 5 refutes this prediction.

Generalizing the Models to Theories

The neoclassical, Keynesian, and classical models that we have presented in this chapter show difficulties in explaining some of the basic facts of capitalism. However, these models are certainly very elementary. One would tend to believe that other models that are constructed under different auxiliary assumptions, such as different markets structures or assuming an open economy, and banking industries, would be needed in order to draw more definitive conclusions about the explanatory power of these theories. As the alpha-beta method rules: to refute a theory requires failure of all of its models.

However, it is relatively easy to see that *any* neoclassical model will predict labor market equilibrium with full employment; it will also predict that money is neutral. The reason is that any neoclassical model will assume the same alpha propositions, which include the assumption that markets are Walrasian. An implication of this primary assumption, *given* the others, is that nominal prices will be flexible and will play the role of clearing the markets, including here the nominal wage rates in labor markets. The other implication is the existence of a dichotomy in general equilibrium: endogenous real variables will be determined by exogenous real variables alone, whereas exogenous nominal variables will only affect nominal variables.

How does neoclassical theory explain the existence and persistence of unemployment? There are two answers in the literature. First, the observed unemployment is mostly *voluntary*; it is the result of worker's choice. As Robert Lucas's (1978) famous statement goes: "The unemployed worker at any time can always find some job at once" (p.354). Where? Supposedly in other labor markets where wages are lower, which then by necessity would have to operate with excess labor demand. It follows that it is logically inconsistent to assume Walrasian labor markets and at the same time observe voluntary unemployment.

Second, unemployment is mostly *frictional*. The model presented above has assumed that workers and jobs are homogeneous. Suppose now that workers and jobs are heterogeneous, which implies that worker and job matching will take time; also suppose that there exists a rate of job separations by firms in the period of analysis. The job-matching process implies that at any period of time there would be workers seeking jobs and also firms seeking workers. Thus, unemployment would be the result of the process of job searching by workers under imperfect information. Then it follows that, given the conditions of demand for labor and supply of labor, there would exist a Walrasian wage rate in the labor market; but the job-matching process would lead to the existence of a "natural rate of unemployment," at which the rate of finding jobs is equal to the rate of job separations (Barro 1997, Chapter 10).

This argument is also logically inconsistent with the assumption of Walrasian labor market, in which the nature of the market is such that the nominal wage rate would change until the market clears. However, under equilibrium with frictional unemployment, the nominal wage rate plays no role in the equilibrium adjustment, which occurs via quantity adjustments. Frictional unemployment would be consistent with the Walrasian market assumption if the mechanism of quantity adjustment would converge toward equilibrium with zero (or negligible) unemployment, that is, as the economic process is repeated period after period, frictional unemployment would have to decline over time toward zero, not toward a positive natural rate of unemployment. The existence and persistence of unemployment simply contradicts the theory: the labor market does not operate as a Walrasian market. Neoclassical theory cannot explain the existence and persistence of unemployment.

Similarly, the conclusion of the particular classical model shown here can be extended to the entire family of short-run models of the classical theory. The reason is that *any* classical model will assume an exogenous real wage rate (an alpha proposition). Therefore, in the short run, unemployment may or may not exist, depending on the factor endowments of society. Any

classical short-run model will also predict money neutrality, which is derived from the assumption of exogenous real wage rate.

In the case of the Keynesian model, it is also simple to show that its results can be generalized to the entire family of models. *Any* Keynesian model will assume "exogenously" determined nominal wage rates, which is however set to generate excess labor supply. Therefore, any model will predict both equilibrium with unemployment and also equilibrium with full employment. In fact, Keynesian theory was created to reach full employment through public policies. Hence, unemployment plays no role in the functioning of capitalism, for it could operate with full employment as well.

The basic conclusion of this chapter is that, regarding the functioning of the First World, Fact (1) refutes the three theories, whereas Fact (4) refutes neoclassical and classical theories. As the alpha-beta method rules: to reject a theory, it is sufficient that the theory fails to explain a single empirical regularity. The problem before us is that an economic theory that seeks to explain the existence and persistence of unemployment—to explain why capitalism *always* operates with unemployment—needs to show that unemployment plays a role in the functioning of capitalism, both in the short run and in the long run. A new economic theory of the First World that seeks to explain both Fact 1 and Fact 4 will be presented in the next chapter, which initiates the construction of partial theories of capitalism leading to a unified theory.

PART II

The Short Run: Production, Employment, and Distribution

CHAPTER 4

The Epsilon Society

Any economic theory that attempts to explain production and distribution in the First World countries must construct an abstract society in which two outcomes of the economic process are necessary. First, the labor market must *always* operate with unemployment. Second, there must be interplay between real and monetary variables. These constitute notable empirical regularities of the First World countries (Facts 1 and 4, as listed in chapter 2). This chapter presents a new abstract society—called epsilon society—with the intention of meeting those challenges.

Epsilon: Socially Homogeneous Class Society and Underpopulated

Epsilon is an abstract capitalist society. It is a class society, in which two social classes exist: capitalists and workers. This class division of society is the result of the unequal distribution of economic assets among individuals. Capitalists concentrate the property of physical capital.

Another assumption about the initial conditions of epsilon society refers to its factor endowments, that is, capital per worker. Capital per worker is such that the marginal productivity of the total labor force is largely positive. Thus, the assumption is that epsilon is an *underpopulated* society. The implication is that the level of labor productivity in society is high enough to allow firms obtain profits after paying wages.

On the social institutions under which the interactions between social classes operate in epsilon, the assumption is that the market system and the democratic political system constitute the basic institutions of capitalism in general and of epsilon in particular. Exchange of goods takes place under

the norms of market exchange, which includes the norms of private property of capital and voluntary exchange of goods and services. The democratic political system includes the norm that individuals are entitled to the same political rights and duties; thus, epsilon is qualitatively a socially homogeneous society.

Labor market institutions are also established. Labor exchange operates under the norms of market exchange, which includes free labor and voluntary exchange (any form of command or slavery is not allowed). Another norm is that market nominal wage rates cannot fall, for this fall would be socially unfair. Thus, reduction of real wage rates may occur only via increases in the price level. The nominal wage rate may go up if firms find it profitable to do so; its stickiness is downward, not upward.

Individuals act motivated by self-interest. However, workers and capitalists follow different objectives. Capitalists want to extract as much effort as possible from workers in order to maximize profits. Workers want to earn the highest real wage with the lowest effort. This social conflict in labor relations arises because of workers' exclusion from the property of the firm; hence, capitalists and workers are not partners.

Another primary assumption is that capitalists seek two objectives that are hierarchically ordered. Capitalists seek, first, to remain as members of the capitalist class and, second, profit maximization. Therefore, secure property rights has priority over profits, that is, there is no substitution between profits and the risk of losing capital endowments.

Extreme degrees of income inequality are not socially tolerated. Individuals have a sense of justice or fairness with respect to economic inequality, which implies a limited tolerance for inequality. Whenever inequality reaches a level beyond the tolerance threshold, individuals will react and seek (legally or illegally) to restore inequality to the tolerable values.

Finally, on organizations, the social institution includes firms, households, and the government. The government is just another social actor that interacts with capitalists and workers; it is not an actor that is above social classes.

The primary assumptions of the epsilon theory (ε) can be summarized as a set of alpha propositions as follows:

$\alpha(\varepsilon)$. *(1) Institutional context*: (a) Rules: People participating in the economic process are endowed with economic and political assets. Economic assets are subject to private property rights. People exchange goods subject to the norms of market exchange, which include the particular norm that in labor markets nominal wage rates cannot fall; the market system operates with Walrasian and non-Walrasian

markets, in which the labor market is of the latter type. The political regime is democratic. (b) Organizations: households, firms, and the government.

$\alpha(\varepsilon)$. *(2) Initial conditions*: (a) Initial inequality: individuals are endowed with unequal quantities of economic assets, but equal political entitlements; hence, there are two social classes: capitalists and workers. (b) Initial factor endowments: the stock of capital per worker is such that the marginal productivity of the entire labor force is largely positive.

$\alpha(\varepsilon)$. *(3) Economic rationality of agents*: Consistent with the institutional context of capitalism, individuals act motivated by self-interest. Capitalists seek two particular objectives, hierarchically ordered: first, maintenance of class position and, second, maximization of profits. In the labor market, workers seek to maximize wages and minimize effort, while capitalists seek to minimize wages and maximize effort. Due to this conflict in labor relations, capitalists use unemployment as the device to extract effort from workers. Individuals have a limited tolerance for inequality.

The Nature of the Labor Market

Could the labor market in epsilon society operate as a Walrasian market? If all workers willing to exchange labor were able to find jobs at the prevailing wage rate, would there be incentives for workers to supply their highest work effort to the firm? What if workers are found shirking and are then dismissed? If the labor market is Walrasian, the dismissed workers will always be able to find jobs at the prevailing market wage rate. What would then be the cost of shirking for workers? None. Workers would then have no incentives to provide their highest work effort in the working place.

When workers and capitalists are not full partners, capitalists must use incentive devices to extract work effort from workers. The incentive device must make shirking costly. What device would do that?

Full employment is against the interest of capitalists. Under a full employment situation, unemployment would cease to play its role of a disciplinary device, as Michael Kalecki argued long time ago (Kalecki 1971, Chapter 12). Contributions that are more recent are due to Shapiro and Stiglitz (1984) and Bowles (1985). Inspired by these works, a labor market theory will now be developed.

In order to create incentives for work effort, an individual firm will seek to pay a premium and offer a higher wage rate than the prevailing market wage rate, which we assume is a Walrasian price. Then dismissed workers will suffer a cost, the premium over the Walrasian wage rate. However, the

other firms will also follow similar procedure and thus the average wage rate in the labor market will rise above the Walrasian wage rate. Then the individual firm, acting guided by the motivation of self-interest, will again increase the wage rate, which the other firms will also follow, and so on.

In order to find convergence, assume that the premium rate, as percentage of the market wage rate, is uniform for all firms and it decreases in every round; thus a final wage rate of equilibrium will be reached. The final outcome is that the new market wage rate will lie above the Walrasian price. The equilibrium nominal wage rate implies excess labor supply or unemployment. Then dismissed workers will not move from one firm to another, but most likely will move to unemployment. Shirking behavior will prove costly for the workers. The aggregate incentive device to discipline workers is unemployment.

Unemployment is thus the device used by capitalist firms to secure labor discipline, which implies maintaining labor productivity at high levels, which in turn implies high profits. Unemployment now plays a significant role in the functioning of the labor market. Assume now that a minimum unemployment rate (call it u^*) is needed to extract work effort from workers. The observed unemployment rate could be higher, due to other factors.

If there must be unemployment in the labor market, then the wage rate cannot be equal to the Walrasian price, but it will have to be higher than that. Moreover, given the labor market conditions of demand and supply, there is a relation between real wage rate and unemployment rate. When unemployment is zero, the market price of labor is equal to the Walrasian price; therefore, a given unemployment rate implies a real wage that is above the Walrasian, and vice-versa. The wage rate that corresponds to the minimum unemployment rate to secure labor discipline may be called *the minimum efficiency wage rate* (w^*). Therefore, wage rates in the labor market must be equal to or higher than the minimum efficiency wage rate as a device to secure unemployment and, thus, labor discipline, which will in turn ensure high labor productivity levels and high profit levels for the firms. The term "efficiency wages" alludes to these properties.

The unemployed would be willing to be hired at the prevailing market wage or even at a lower wage rate; nonetheless, they cannot bid the wage rate down as a way to get employed. Firms would not accept that because they know that if they hired workers at lower wage rates than w^*, work effort and labor productivity would decrease. Thus, no one has the power and the incentive to change this solution and there is equilibrium with unemployment in the labor market. The labor market is non-Walrasian. This is a case in which, notwithstanding the endogenous determination of the real wage rate, the labor market is non-Walrasian. The equilibrium real wage is not the

Walrasian price and the labor market equilibrium is necessarily with excess labor supply.

In sum, in epsilon society, the labor market *always* operates with unemployment; it is a non-Walrasian market. A proportion of workers will be excluded from the labor market. The labor market cannot operate as a potato market (the paradigm of a Walrasian market); potatoes cannot change their behavior depending on their market price or depending on what the excess supply is, but workers can. This is the nature of the labor market.

A labor market model of the epsilon theory is now needed to make the theory falsifiable. Introduce explicitly the work intensity factor (λ) in the production function. It is a factor that determines the level of output for a given quantity of labor and capital inputs and for a given length of the working period. For the firm *j*, producing good B, the production function can be written as

$$Q_{bj} = \lambda\Phi_j (D_{hj}, K_{bj}) \tag{4.1}$$

Here D_{hj} is the quantity of labor demanded in the labor market and K_b is the firm's capital endowment.

Assume that the work intensity is endogenous and depends upon the cost that the worker would have to pay if engaged in shirking behavior at the workplace. This cost will depend upon the possibility of being caught shirking, which in turn will depend upon the supervision resources used by the firm (assumed to be a fixed cost) and the income losses the worker would suffer if fired from the job. If there were always full employment, that is, if the labor market were Walrasian, the latter cost for workers would be zero because workers who were found shirking and then dismissed from the firm would always be able to find another job at the prevailing market wage rate; hence, workers would have no incentives to supply their highest work effort.

For workers, the cost of shirking is higher if the unemployment rate (u) is higher. However, given the labor market conditions of supply and demand, the rate of unemployment and the real wage rate are not independent of each other; moreover, there is a one-to-one relationship (and a positive one) between unemployment rates and real wage rates.

In order to simplify the model even further, consider only two levels of effort. Then

$$\lambda = f(u), f' > 0, \text{ such that} \tag{4.2}$$
$$\lambda = \lambda^* = 1 \text{ if } u \geq u^* \text{ or } w \geq w^*$$
$$\lambda = \lambda' < \lambda^* \text{ if } u < u^* \text{ or } w < w^*.$$

Here u^* is the threshold of the unemployment rate above which workers do not shirk (λ^*) and below which they do (λ'). This threshold is socially determined. The degree of partnership in labor relations is assumed to be the essential factor, which is taken as exogenously given. Unemployment is thus a device for labor effort extraction, or labor discipline. The term w^* is the wage rate that is consistent with the unemployment rate u^* in the labor market. It is also a threshold value leading to labor discipline, that is, $w^*$is the minimum efficiency wage rate.

The individual firm now faces two curves of the marginal productivity of labor, one for each level of effort. These can be represented by the following equations:

$$\lambda^*\Phi_{j1}\,(D_{hj},\, K_{bj}) \text{ if } u \geq u^* \text{ or } w \geq w^* \qquad (4.3)$$
$$\lambda'\Phi_{j1}\,(D_{hj},\, K_{bj}) \text{ if } u < u^* \text{ or } w < w^*.$$

Adding horizontally over all firms, the aggregate curve of the marginal productivity of labor is determined.

Figure 4.1 panel (a) shows the labor market model. The curve H*R*N* represents the marginal productivity of labor when the degree of work intensity is equal to λ^*. The curve H′R′N′ represents the marginal productivity of labor when the degree of work intensity is equal to λ'. Because under perfect competition the labor marginal productivity curve also represents the labor demand curve, there seems to be two demand curves, one for each work effort level; but this is not the case. Assume labor supply is exogenously given at the value S_h. Therefore, the upper curve requires that the unemployment rate be equal to or greater than the threshold value u^*; if the unemployment rate is lower than the threshold value, then the relevant curve is the inferior one.

Given the value of u^*, the maximum level of employment (D_h^*) that can maintain the optimal work intensity (λ^*) is determined. This is also shown in figure 4.1 panel (a): the segment H*R* is the relevant marginal productivity of labor when employment is equal to or less than D_h^*; if employment goes beyond D_h^*, the marginal productivity curve jumps down to the curve H′R′N′.

Where is the labor demand curve now? At the wage rate w^*, the quantity of labor demanded would be equal to D_h^*. At levels of real wage above w^*, the quantities of labor demanded are indicated by the segment H*R* of the upper marginal productivity curve. At values of real wage below w^*, would the quantity of labor demanded reach the segment R*N*? No, because this would imply that total employment is beyond D_h^*, which in turn would

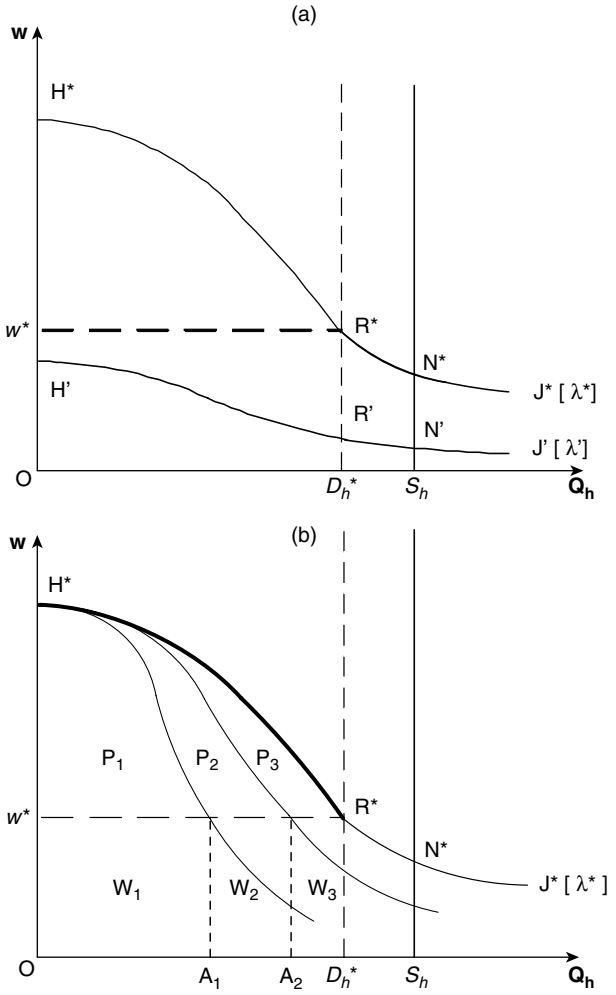


Figure 4.1 Labor productivity and labor demand.

imply that the unemployment rate would take a value below u^* and consequently work intensity would fall and the marginal productivity curve would shift downward to curve H′N , which would lead to a set of smaller profits than at w^*.

Firms would thus have no incentives to pay real wages below w^*. If, for some reason, the market real wage rate were below w^*, firms would

have incentives to pay higher nominal wage rates so as to raise the real wage rate and thus to shift the labor marginal productivity to the upper curve. The real wage w^*operates as a floor price, as a threshold value of the set of efficiency wages. Therefore, of the marginal productivity of labor curve H*R*N*, only the segment H*R* (the thick line) represents the labor demand curve.

It should be noted in figure 4.1 panel (a) that profits could be higher than what they are at point R* if the real wage rate could take a value below w^* but holding constant the total employment level D_h^*. However, this solution would require collective action behavior of capitalists, which is ruled out here due to the selfish and individualistic rationality. The reason rests upon the assumption of the motivation of self-interest, which leads capitalists to free-ride behavior, the so-called Olsonian problem (Olson 1965): an individual firm would seek to hire more workers to make even higher profits, but all firms will follow this behavior, and thus more employment would result in the aggregate. Therefore, the vertical segment $R^*D_h^*$ is not part of the labor demand curve.

According to this model, the labor market cannot operate as a Walrasian market, that is, the solution cannot occur at point N* in figure 4.1 panel (a). It appears that firms can make higher profits at point N* than at point R*, but this is not the case. At a real wage rate below w^*, the marginal productivity of labor curve would shift down to H′R′N′. Point N* is an economic illusion; it does not exist. The Walrasian real wage rate is the one that corresponds to point N′, not to point N*. The assumption of the efficiency wage model is that total profits are higher at point R* than at point N′. The market real wage rate is endogenously determined, but subject to the condition that it must be higher than the Walrasian price.

In sum, the segment H*R* represents the labor demand curve and along this segment the real wage rate will be determined endogenously. *The labor demand curve is not equal to the entire marginal productivity of labor curve, but only to the upper segment of this curve.*

To be sure, price and quantity of equilibrium in the labor market are still determined by the interaction of supply and demand conditions, but subject to the constraint that the market real wage rate must be equal to or higher than the minimum efficiency wage (w^*). This is equivalent to the condition that the unemployment rate must be equal or higher than a minimum rate (u^*). The minimum efficiency rate of unemployment at the same time determines the maximum employment level that is consistent with the optimum work effort of workers, as judged by capitalists, which may be called *the effective full employment level* (D_h^*). Full employment in the sense of zero rate of unemployment cannot be attained in the labor market.

Assume now that firms have different production functions due to the endowment of specific factors, such as entrepreneurial talents and location. The marginal productivity of labor will now be firm specific, even if firms are endowed with the same capital stock. Then the aggregate curve of marginal productivity of labor is the result of ordering the individual curves of firms, from the highest to the lowest levels, and then aggregated them horizontally. This is shown in figure 4.1 panel (b). The aggregated curve is also called H*N*, as in figure 4.1 panel (a), just to show the composition of firms underlying this curve. At the value of real wage w^*, we can see the composition of employment, wage bill, and profits by groups of firms. Profits are the lowest in the group of firms that is endowed with the lowest specific factors. Hence, the aggregate marginal productivity of labor falls due to diminishing returns within firms (more workers working with the same stock of capital) and to differential labor productivity levels among firms.

In conclusion, according to this labor market model of the epsilon theory, the labor market is non-Walrasian; labor market equilibrium is necessarily with unemployment. The labor market model assumes that unemployment is the labor discipline device that firms use to maximize profits. Unemployment is a necessity for the functioning of capitalism in epsilon society. Can the general equilibrium model of the epsilon theory lead to this result?

A Static Model of the Epsilon Theory: The Short Run

In order to make the epsilon theory falsifiable, an epsilon model must be constructed, which can be compared to the task of constructing a particular stage set in which social actors will play their roles. The following auxiliary assumptions are introduced for that purpose:

- There are two social classes: the capitalist class, which includes the owners of physical capital (very small group) and the high executives or managers, who are partners of the capitalists, and the workers, who are endowed with stocks of human capital and cash balances. Government behavior consists of supplying money to the economic process.
- The initial efficiency nominal wage is given and the rule is that it cannot fall. Wages are paid at the end of the production period; consequently, workers need to hold cash balances for transaction purposes. There is no market for renting capital services.
- Four markets constitute the market system: labor, commodity, money, and foreign exchange; all are Walrasian, except the labor market. Human capital is the same for all, capitalists and workers, and

therefore there is only one labor market. The labor market operates with efficiency wages (as shown earlier). One sole good (B) is produced in society. Workers and capitalist firms seek to exchange labor services and goods in the labor and commodity markets; money supplied by the government is used as the means of payment. Capitalist firms seek to exchange with international markets, using foreign exchange as the means of payment.

- Capitalists, workers, and the government interact and compete in the market to achieve their goals, but none has the power to set prices (perfect competition in the market system); they all face costless information on market prices and technology. Labor productivity levels differ by firms due to differences in intrinsic factors, such as entrepreneurship talents and locations.
- The economic process is static and open.

Behavior of Workers

Workers know that the labor market functions with unemployment: not all workers will be able to exchange the quantities of labor they are willing to sell at the prevailing market wage rate. Once firms have determined the total quantity of labor demanded, the quantity of workers that capitalists will hire from each household is exogenously determined through a random mechanism of exclusion.

The behavior of workers will be similar to that shown in the Keynesian model in the previous chapter. The reason is that workers will face in epsilon society the same context as in that model. As for rationality, it is assumed that individual workers seek to maximize total consumption subject to the required real cash balance constraint and the real income constraint. Just for convenience, the structural equations for worker i are repeated here:

$$P_b D_{bi} = P_b D_{bi} + (D_{mi} - S_{mi}) \qquad (4.4)$$
$$D_{mi} = P_b L_i (D_{bi}),\ L_i' > 0,\ L_i'' < 0.$$

Note that workers cannot decide on the quantity of labor services that they want to sell in the labor market because the labor market is non-Walrasian. Given the values of the exogenous variables, workers will choose the values of the endogenous variables (quantity of good B to buy and quantity of cash balances to hold) according to this rationality. The equilibrium condition is that the cash balances be willingly held ($S_{mi} - D_{mi} = 0$), which implies that total income is spent in buying consumption good B.

The static equilibrium is (trivially) stable. Therefore, the comparative statics method can be applied to derive beta propositions. Thus, the individual worker's demand for good B and demand for money can be derived from equation (4.4) as follows:

$$D_{bi} = F^i (P_b, D_h, P_h),\ F_1<0,\ F_2>0,\ F_3>0 \qquad (4.5)$$
$$D_{mi} = P_b L_i (D_{bi}),\ L'>0,\ L''<0.$$

Behavior of Firms

The model assumes that the epsilon economy is small in the international economy; so international prices are exogenously given. Epsilon economy exports good B and imports good C, which is a material input required to produce good B. Hence, the domestic price of commodities B and C are equal to

$$P_b = P_e P_b^* \qquad (4.6)$$
$$P_c = P_e P_c^*.$$

P_e is the nominal exchange rate (the domestic nominal price of foreign exchange) and P_b^* and P_c^* are international prices denominated in units of the foreign exchange. Each domestic price must be equal to the international price multiplied by the nominal exchange rate. If this equality did not hold, people could buy commodities from the cheaper source and sell it wherever it is more expensive and make a profit. The ratio $z^* = P_b^*/P_c^*$ is called the *international terms of trade*.

It should be clear that the domestic price level P_b refers to the good produced domestically (good B), which depends upon the exchange rate and the international price of good B. Given the value of the latter, the domestic price level varies according to the variation of the exchange rate. Price stability implies exchange rate stability.

Firms seek profit maximization. Nominal profits is equal to the difference between total revenues minus total costs; total revenues is equal to price times quantity of good B sold in the market; total costs is equal to the sum of the wage bill (nominal wage times quantity of workers hired) and the total cost of material inputs (domestic price times quantities of good C imported).

Production technology is such that output of good B depends on the services of the stock of physical capital endowed by each firm, the quantity of workers hired, and the material input C bought abroad. Assume that

input C enters into the production of good B in fixed quantities per unit of output, that is, assume that input C is not substitutable for capital or labor, but capital and labor are substitutable to each other. Then for every unit of output, the firm must pay a given amount for input C, which is required for technological reasons for the production of good B. Therefore, the *net* value of average labor productivity is now the relevant concept, which is equal to total net value of output (net of the cost of input C) divided by the number of workers; similarly the *net* value of marginal productivity of labor is the relevant concept, which is equal to the additional *net* value of output that results from hiring an additional worker.

The behavior of the individual firm *j* can then be represented by the following structural equations:

$$\text{Maximize } P_j' = P_b\, Q_{bj} - P_h\, D_{hj} - P_c\, D_{cj} \tag{4.7}$$

subject to the production function constraints,

$$Q_{bj} = \lambda\Phi_j\,(D_{hj},\, K_{bj}), \text{ such that } \lambda = \lambda^* = 1$$
$$D_{cj} = c\, Q_{bj}$$
$$X_{bj} = (1/z^*)\, D_{cj,}$$

and subject to efficiency wages, that is,

$$\lambda = \lambda^* = 1 \text{ requires that } w \geq w^* > w^w.$$

Profits (P') refer to nominal values. Total output is net of depreciation, so is total profits. The term c is a technological coefficient, equal to the number of units of good C required to produce one unit of good B. The two first constraints assume that the production function is subject to *limitational factors* and is thus represented as a system of equations.[1] The third constraint answers the question: Who does the firm pay for the imported input C? Because good C is an imported input, the firm pays foreign firms with foreign exchange that obtains by exporting part of its output B (X_b) according to the terms of international trade z^*. The last constraint indicates that efficiency wages must prevail in the labor market, which must be higher than the Walrasian real wage rate (w^w). Technology level in the production of good B and the production skills embodied in workers, known as *human capital*, will be held constant in the short run and for simplicity are just ignored in the production function.

The structural equations presented above lead to the following equilibrium condition:

$$\Phi_{j1}\,(D_{hj},\, K_{bj})\,(P_b - c\,P_c) = P_h \text{ or} \qquad (4.8)$$
$$\Phi_{j1}\,(D_{hj},\, K_{bj})\,(1 - c/z^*) = w, \text{ where } (c/z^*) < 1.$$

The marginal productivity of labor is now net of the cost of using the limitational input C. Profit maximization implies the equality between the value of the net marginal productivity of labor and the nominal wage rate or, alternatively, the equality between the net physical marginal productivity of labor and the real wage rate. Profits arise from the difference between the average productivity of labor and the marginal productivity of labor multiplied by the employment level.[2]

Individual firms are price takers in all markets. Therefore, for each firm, profit maximization implies hiring workers up to the point that the value of the net marginal productivity of labor is equal to the market nominal wage rate. Given the exogenous variables, firms will choose to produce a quantity of good B, buying a quantity of input C and hiring a quantity of workers. As long as the exogenous variables remain fixed, these quantities of the endogenous variables will be repeated period after period. This is the static equilibrium in the behavior of the typical firm.

The equilibrium situation is (trivially) stable. Comparative statics can then be applied to determine beta propositions about the behavior of the firm. The critical variables are the quantity of labor demanded and the quantity of good B supplied to the domestic market. They are

$$D_{hj} = H^j\,(P_b,\, P_c,\, P_h,\, K_b),\; H^j_1 > 0,\; H^j_2 < 0,\; H^j_3 < 0,\; H^j_4 > 0 \qquad (4.9)$$

subject to

$$P_b\,S_{bj} = P_b\,Q_{bj} - P_b\,X_{bj} - P'_j = (P_c\,D_{cj} - P_b\,X_{bj}) + P_h\,D_{hj.}$$

At the firm level, the quantity of labor demanded depends positively on both the price level and the capital stock because an increase in the value of these variables increases the value of the marginal productivity of labor. It also depends negatively on the price of the input C because an increase in this price reduces the net value of the marginal productivity of labor. Finally, it depends negatively on the nominal wage rate because an increase in this price would lead the firm to increase also the value of the marginal productivity of labor, which in turn would imply a fall in employment (due to diminishing returns). The second equation is the budget constraint of the firm, which shows that what the firm sells in the market must allow the firm to pay for the purchase of the production

factors. The quantity supplied of good B to the domestic market is derived from this equation.

Government Behavior

As any other social actor in epsilon society, a politician will also act motivated by self-interest. Politicians are not to be considered special people whose actions are motivated by altruism, for that will be inconsistent with the social context of an epsilon society. Politicians also constitute a social class. In representative democratic systems, citizens elect governments and empower them. This is the origin of political power. Therefore, politicians seek a place in government by investing in the electoral competition; once in government, they seek to maintain political power and incomes. They give priority to the objective of maintaining their position in the political class. The theory of government behavior is thus much alike the theory of capitalist behavior.

This rationality of political actors also implies a hierarchy on government policies. Policies that have effects in the short run (and influence the voting in the next elections) have priority over those that have long-run effects (when the influence on the next elections is not significant). Government behavior will then tend to be myopic and shortsighted to address the long-run problems of society.

Auxiliary assumptions will be introduced now to construct a simple model of the government theory. Assume that governments seek to maximize votes, subject to the following constraints: the public budget, the force of pressure groups upon the public budget, and the mandatory expenses associated with the financing of rights. On the government budget, the assumption is to consider money supply as the only source of government revenue. The government prints money and transfers it to individuals in the lump sum form, as if by helicopters. This mechanism is known as "helicopter money." Other income sources such as taxes and debt will be ignored in this model.

Now suppose that voting for the current government depends positively upon the indicators of the economic process. Higher national income level (which implies higher employment level and lower unemployment level), higher real wages (or lower inflation rate) will increase the voting support to the government. The only instrument the government has is the supply of money. The government will thus respond to the outcomes of the economic process with changes in the quantity of money according to its own interest. The political cycle will also change the government expenditure. The government will seek to attract votes by increasing the money supply in electoral periods.

In sum, government behavior is also motivated by self-interest; therefore, the supply of money will be endogenous. The government is not above class relations, but rather, interacts with the other social classes following its own interests. Therefore, its social function is accomplished as a by-product of its selfish motivation. This government theory was inspired by the work of Downs (1957). (A more elaborate model of government behavior will be developed in chapter 7.)

General Equilibrium

It is now time to aggregate the individual behavior of social actors. Aggregating over all workers, the market demand functions for good B and for money are determined. This is simply the result of adding all individual functions, for workers make their choices independently, not by group interactions or collective actions.

The market behavior of workers is presented as the following structural equations:

$$D_b = F(P_b, D_h, P_h),\ F_1<0,\ F_2>0,\ F_3>0 \qquad (4.10)$$
$$D_m = P_b\, L(W),\ L'>0,\ L''<0$$

subject to the aggregate budget constraint of workers

$$P_h\, D_h = P_b\, D_b + (D_m - S_m).$$

The quantity of good B demanded depends negatively on the price level, positively on the nominal wage rate, and positively on the level of employment of the economy, that is, it depends on the quantity of workers that firms will hire, not on the quantity of labor that workers are willing to sell. The higher the employment level, the higher the quantity of good B demanded will be. (In a Walrasian market, where workers sell all their labor supply, the quantity demanded would depend upon their total income, which in turn depends on nominal wages only.) The demand for money depends on the price level and the wage bill because the demand comes from workers alone.

Aggregating over all firms, the market behavior of the firms can be written as the following set of structural equations:

$$D_h = H(P_b, P_c, P_h, K_b),\ H_1>0,\ H_2<0,\ H_3<0,\ H_4>0 \qquad (4.11)$$

subject to

$$P_b S_b = (P_c D_c - P_b X_b) + P_b D_b$$
$$X_b = (1/z^*)\, D_c = (1/z^*)\, f(D_b, K_b), f_i > 0$$
$$w \geq w^*.$$

The first equation is the labor demand function. The second equation is a constraint and shows the quantity of good B supplied to the domestic market, which is derived from the firms' budget equations. The third equation is also a constraint and represents the condition that firms must pay for the import of input C to foreign firms, that is, they must export good B for the amount X_b. This is also the foreign exchange market equilibrium condition. The last constraint refers to efficiency wages.

From the structural equations shown above, the general equilibrium conditions can now be established. They are

$$\text{Labor market } S_b \equiv D_b + U,\ P_b \geq P_b^*,\ w \geq w^* \qquad (4.12)$$
$$\text{Good B market } S_b = D_b$$
$$\text{Money market } S_m = D_m$$
$$\text{Foreign exchange } P_b^*\ X_b = P_c^*\ D_c$$

subject to the aggregate budget constraint

$$P_b (D_b - S_b) + P_e (P_c^*\ D_c - P_b^*\ X_b) + (D_m - S_m) = 0.$$

The labor market is non-Walrasian, but the rest are Walrasian. The equilibrium condition in the good B market refers to the domestic market, in which the demand comes from workers only; the equilibrium condition in the money market states that workers must willingly hold the quantity of money supplied by the government; and the equilibrium condition of the foreign exchange market comes from the budget equation of firms.

From the aggregate budget constraint of system (4.12), Walras's law is derived: one of the three Walrasian markets is redundant. Thus, the system needs to solve only for three markets: two Walrasian and the non-Walrasian.

The general equilibrium must solve for the endogenous variables of the system for given values of the exogenous variables. These variables can be separated as

$$\text{Endogenous variables: } P_e, D_b, P_b, P_c, D_b, S_b, D_c, X_b, U, Y, D$$
$$\text{Exogenous variables: } z^*, K_b, S_b, \delta, S_m$$

Total output (Y) and its distribution (D) are endogenous. Money supply (S_m) appears as exogenous, just for operational convenience, but it belongs to the special category of *quasi-exogenous*, as will be argued next. The initial degree of inequality in the distribution of economic assets (δ) appears explicitly in the set of exogenous variables.

Notice that some of the endogenous variables at the general equilibrium level of analysis were exogenous at the microeconomic level, such as the nominal prices. This is in accord with the principle of increasing endogenization of variables in the aggregation process. In this general equilibrium model, total output and distribution of equilibrium are the result of the interactions between the three social actors of the model: capitalists, workers, and the government. Just for the sake of simplicity, the general equilibrium solution will be solved sequentially: given the quantity supplied of money, the market system determines total output and its distribution; then the government acts endogenously to modify the market result by changing money supply so as to achieve another general equilibrium with higher output level and higher real wage rate, subject to the constraint of having price level stability and the demands of the pressure groups. Thus the quantity of money is said to be quasi-exogenous in this analytical sense.

Now the task is to identify the core of the general equilibrium system. Start with foreign exchange market. Given that this is a one-good economy, in which imports are production inputs (not for direct consumption), the condition of equilibrium is very simple: firms must export good B to pay to foreign firms for the imported input C. The higher the exchange rate, the higher the price level, and the higher the quantity of labor demanded and also the quantity of input C, which implies a higher quantity of foreign exchange demanded; therefore, the quantity of foreign exchange demanded will increase with total output. However, the quantity supplied of foreign exchange will always be equal to the quantity demanded because firms import and export at the same time and thus decide for the quantity demanded and supplied of foreign exchange. The nominal exchange rate will be determined in the money market.[3]

From the labor demand function, it follows that the labor demand curve has the property that at each nominal wage rate there will exist a quantity of labor demanded, which also determines the quantity of input C and therefore the quantity of foreign exchange demanded. Since firms must directly pay for input C by exporting part of output B, the quantity of foreign exchange demanded must be equal to the quantity supplied. Therefore, equilibrium in the foreign exchange market will be ensured at each employment level. This property is shown by the second equation among the constraints in the system (4.11), which constitutes the equilibrium condition

of the foreign exchange market. Thus, the equilibrium condition in the foreign exchange market can be incorporated into the labor market. The assumptions of input C as a limitational factor and one-good economy enable us to do this.

The core of the general equilibrium system is then constituted by two markets alone: the labor and money markets. The core system can be written as follows:

$$D_h = H(P_e; P_h, z^*, K_b),\ H_1>0,\ H_2<0,\ H_3>0,\ H_4>0 \qquad (4.13)$$
$$S_m = P_e\ P_b^*\ L(W) = M(P_e, D_h; P_h,\ P_b^*),\ \text{where } M_i > 0,\ \text{for all } i$$

subject to $w \geq w^*$.

The first equation is the labor demand function; the second is the equilibrium condition in the money market. These two equations solve for the two endogenous variables, the employment level and the nominal exchange rate. The constraints indicate that we are searching for the existence of general equilibrium with efficiency wages, which implies equilibrium with unemployment.

Figure 4.2 displays the core of the general equilibrium model of the epsilon theory. The labor market in terms of the nominal wage rate is presented in panel (a). The value of the gross marginal productivity of labor is represented by the curve h′n′ and the value of the net marginal productivity of labor by the curve hn. The gap between these two curves measures the cost of using the limitational factor C, which is paid with exports of good B. Given the nominal wage rate, the labor market will determine the employment level, which must not exceed the level D_h^*.

The money market equilibrium can be established using the exchange rate instead of the price level, as shown in panel (b). The higher the exchange rate, the higher the domestic price level, and, thus, the higher the demand for nominal money. A downward sloping demand curve for money can be obtained by using $1/P_e$ as the "price of domestic money in units of the foreign exchange." Given an initial stock of money, there will be an exchange rate value that clears the money market.

It follows from the graph that the solution of the core is simultaneous. The labor market must solve for the employment level, which requires knowing the foreign exchange value. The money market must solve for the foreign exchange rate, which requires knowing the employment level. There is one, and only one, set of values of the exchange rate and the employment level that can produce equilibrium in both markets. Point e and point F show the core solution in figure 4.2.

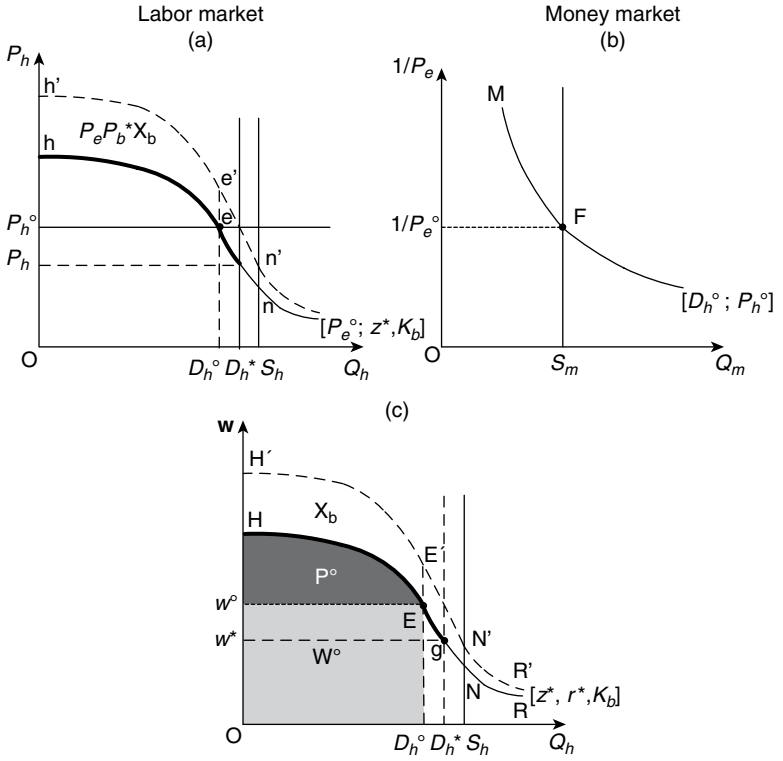


Figure 4.2 General equilibrium in epsilon society.

From the solution of the core, the equilibrium values of the rest of the endogenous variables can be found by implication only. Therefore, once the exchange rate is known, the price level is also determined. Then, the other real variables, such as the real wage rate, are also determined.

The labor demand curve related to the real wage rate is presented in panel (c). The curve H′R′ represents the gross marginal productivity of labor and the curve HR the net marginal productivity of labor. The gap between these two curves measures the quantity of exports of good B needed to pay for the use of the required quantity of imported input (intermediate good C). Therefore, at each employment level there is equilibrium in the foreign exchange market; moreover, the labor demand curve now has another parameter: the international terms of trade z^*.

Because the work intensity level depends on the unemployment rate, panel (c) also shows that there is a minimum unemployment rate (u^*) that

assures this intensity. Given the labor supply, u^* implies a minimum wage (w^*) for the set of efficiency wages and also the maximum employment level (D_h^*); therefore, *only the segment Hg (of the curve HR, the marginal productivity of labor) represents the labor demand curve*. If for some reason, the real wage rate took the value at full employment, at point N, the curve HR would shift downward; hence, point N is unattainable.

The real wage rate is consistent with the employment level of equilibrium found in the core. In panel (c), equilibrium occurs at point E. The area under the curve H′E′ measures total output and the area under the curve HE total output net of exports. The gap measures the quantity of good B exported to pay for the required quantity of input C in producing the gross output. Thus, the foreign exchange market is also in equilibrium.

The area under the curve HE (not under H′E′) measures total net output, net of depreciation and net of the cost of imported inputs. This is distributed among workers (the wage bill W) and capitalists (profits P). This is the income distribution of equilibrium. The commodity market should also be in equilibrium due to Walras's law, which can be confirmed graphically: the area showing total quantity of commodity B supplied to the domestic market (equal to total net output minus profits retained by capitalists) is equal to the area of the total wage bill, which is equal to the quantity demanded for commodity B domestically. The market demand for good B comes only from wage earners.

Production and distribution of equilibrium have thus been solved. The total output of equilibrium is called the national income (Y). The distribution of this income between social classes, as total profits and total wages, can be written as follows:

$$Y = P + W = P + w\,D_h \qquad (4.14)$$
$$S_h = D_h + U$$

In order to make it clear that the capitalist system operates with unemployment, the labor allocation equation includes the size of unemployment (U) with zero wage income.

To have social viability in the general equilibrium, the model assumes that the unemployed get incomes from the public budget as unemployment insurance, part of the economic rights established in epsilon society. To maintain the same labor productivity level, the transfer per worker will have to be smaller than the market wage rate.

The general equilibrium solution is indeed an equilibrium situation because no one has both the power and the incentive to change it. The

values of the endogenous variables will be repeated period after period as long as the values of the exogenous variables remain fixed.

As stated before, this is a theory about the market system behavior, which assumes that the market system is able to solve this problem. Certainly, the theory does not say that markets solve for prices and quantities by actually solving equations; it rather says that whatever mechanism markets have to solve this problem (auctioneers, telecommunications, brokers, trial and error, etc.), it is equivalent to solving equations. The theory says that the market system operates *as if* it solved a system of equations, as if it were a big computer.

A comparison of these results with those of standard economics is in order. Macroeconomics textbooks consider models in which total unemployment is analytically decomposed into excess labor supply and frictional unemployment (Barro 1997; Krugman and Wells 2006). Frictional unemployment is the result of job-matching problems: because workers and jobs differ, workers search for jobs and firms search for workers, at the given Walrasian wage rate in the labor market. In this epsilon model, by comparison, frictional unemployment will be ignored because all workers and all jobs are alike and job search problems could hardly exist or will be very small; thus even in that situation, general equilibrium with unemployment is the outcome. Unemployment as excess labor supply is thus a structural characteristic, a necessity, for the functioning of the epsilon society.

The assumption of a minimum unemployment rate made in the labor market model of the epsilon theory has also been introduced in some macroeconomics textbooks, from which the "natural rate of unemployment"—now defined as the nonincreasing inflation rate of unemployment—has been derived (Blanchard 2009, Chapters 6 and 8). The necessary unemployment rate (u^*) of the epsilon model is thus equivalent to the natural rate of unemployment in the sense that in both cases an increase in money supply (or in any other nominal variable) will increase the rate of inflation without reducing further the rate of unemployment.

Beta Propositions

The epsilon model has been constructed in such a way that it is a logically correct system. As we know, a logically correct theoretical model may be empirically false because the model is based on arbitrary assumptions. Falsification of the model is therefore mandatory, which implies confronting the beta propositions of the model with facts. It is relatively easy to show that the core of the general equilibrium is stable. Stability is trivial in the labor market (a non-Walrasian market), as firms will be able to readjust

automatically if the quantity of employment is for some reason out of equilibrium, that is, there cannot exist instability in the labor market. Therefore, the only place where instability would possibly exist is the money market (a Walrasian market). The money market is indeed stable because the demand curve slopes downward and the supply curve is a vertical line; hence, any price that is for some reason out of equilibrium would restore its equilibrium value automatically. In this very simple model, stability of the general equilibrium requires stability in the money market only, which is fulfilled. Therefore, comparative statics can be applied to the general equilibrium solution to derive beta propositions.

The effect of changes in the exogenous variables upon the endogenous variables in the short run, when the productive capacity of society is given, can now be analyzed. The relevant exogenous variable include the international terms of trade (z^*) and money supply (S_m).

Two equilibrium situations must then be analyzed, one in which unemployment rate is above the necessary rate ($u>u^*$) and the other in which it is equal to it ($u=u^*$). Full employment is usually defined as excess labor supply equal to zero ($D_h=S_h$). This is unattainable in epsilon society. What is relevant is the *effective full employment* ($D_h^*= (1-u^*)S_h$), which is necessarily smaller than the quantity supplied of workers.

Consider an increase in the international terms of trade (z^*). The initial equilibrium is showed at point E in figure 4.2 panel (c), where $u>u^*$. The comparative statics shows that the new equilibrium will imply higher real wage rate and higher employment level. The mathematical proof is presented in the Appendix.

An intuitive adjustment can however be indicated as follows. The effect of an increase in variable z^* is to shift the labor demand curve outward. At the same employment level, the total net output will be higher, as exports to pay for imports will be smaller. Firms would now seek to hire more workers at the current real wage rate from the pool of the unemployed; hence, the wage bill would increase, which would imply a higher quantity of real cash balances demanded; but given the quantity of money supplied, the buildup of cash balances would imply a temporary fall in the quantity of good B demanded, which would increase both exports and the inflow of foreign exchange, leading to a fall in the exchange rate in the money market. Then there would be a fall in the domestic price level and a rise in the real wage rate; hence, employment would tend to fall, but not to offset the initial increase.

Just for simplicity, suppose the new wage bill remains unchanged with this adjustment, which implies no further change in the real cash balance. Thus, the new equilibrium will be reached: with higher real wage rate and

higher employment level. The initial equilibrium situation, given by point E, will be moved to another equilibrium situation, say point E″ (not marked), which will be located to the northeast of point E.

The effect of an increase in the international terms of trade upon income inequality can also be seen in figure 4.2 panel (c). The new equilibrium implies higher output level, higher wage bill, and higher profits; hence, the change in income inequality will be ambiguous.

Consider now the effects of an increase in the money supply. It would generate an excess of cash balances willingly held by workers, who will then intend to restore the equilibrium by using this excess to buy more quantities of good B in the period of adjustment. This adjustment would reduce both the exports and foreign exchange inflows and thus an increase in the exchange rate. Then the price level would increase and consequently the real wage would fall; hence, firms would seek to hire more workers and produce more output to gain more profits. This would be the first-round effect.

The second-round effect refers to the feedback effects. A fall in the real wage rate and an increase in the employment level imply an ambiguous change upon the real wage bill and thus upon the demand for real cash balances and upon the price level. Therefore, the second-round effect upon the price level will be small and may safely be neglected. In sum, the effect of an increase in money supply is to decrease the real wage rate and to increase the employment level. The mathematical proof is presented in the Appendix.

The effects of an expansionary monetary policy on output and its distribution can also be seen in figure 4.2 panel (c). Point E represents the initial general equilibrium situation. An increase in the money supply increases the price level, which implies a fall in the real wage rate. The employment level increases and total output also goes up. The initial equilibrium situation, at point E, will be moved to another equilibrium situation, which will lie below point E (not marked), along curve HN. The government can then reduce unemployment using monetary policy, even though at the cost of increasing the price level and reducing the real wage rate. However, the objective of maximization of votes may create incentives for reducing unemployment even at that cost because unemployment is a risk all workers face.

A sufficient large increase in money supply will ultimately reach the effective full employment level. The equilibrium situation can move from point E to point g in figure 4.2 panel (c). If money supply keeps increasing, employment will remain fixed, and only the exchange rate and, consequently, the price level will rise, which will then induce increases in the nominal wage rate in the same proportion in order to avoid the fall in the minimum efficiency real wage rate (w^*), that is, money becomes neutral at

the effective full employment level. It should be clear that the Walrasian market equilibrium, point N, is unattainable in the epsilon model.

On income distribution, the effect of increasing the money supply is to increase profits, whereas change in wage bill is undetermined. About changes in personal income inequalities: profits increase and real wage falls, but the proportion of unemployment falls, which implies an undetermined change in income inequality. Under effective full employment, when money is neutral, income inequality will remain unchanged.

The epsilon model allows for the supply of money to become endogenous when the government chooses to react to the results of the market system, seeking to alter these results in the direction indicated by the logic of maximization of votes. Money supply is the only instrument that the government can manage to achieve this objective. Money supply is not purely exogenous or endogenous; hence, it could be classified as a quasi-exogenous variable in this model.

For example, consider an external shock, such as a significant fall in the international terms of trade. We already know that the effect will be a fall in both total output and employment. Suppose the government seeks to react to this shock with monetary policy to offset the negative effect. This government behavior implies that the *net effect* of the exogenous variable international terms of trade upon endogenous variables include the direct effect and the induced effect upon money supply, which can be written as follows:

$$Y = F(z^*, S_m),\ F_i > 0 \tag{4.15}$$
$$S_m = H(z^*),\ H' < 0$$
$$Y = F(z^*, H(z^*)),\ \text{such that}\ \partial Y/\partial z^* = F_1 + F_2 H' > 0.$$

The first equation indicates that, in the short run, the equilibrium level of output Y depends positively upon the terms of international trade and the quantity of money. The second equation says that government behavior responds to this external shock by changing the money supply. Vote maximization behavior would lead the government to have countercyclical monetary policies: An external negative shock causes economic recession, which induces the government to increase money supply to counteract this shock.

The third equation assumes that the external shock cannot be offset totally with monetary policy. The model of vote maximization behavior predicts that government policy would seek to maintain effective full employment, that is, limit unemployment to only the necessary unemployment rate. Nevertheless, the money supply effect includes increases in the price level as the effective full employment situation is approached, which may

imply increases in the inflation rate. Votes may decline as a result. Hence, vote maximization behavior will stop short of effective full employment.

In the epsilon model, the government has two policy instruments: money supply and the nominal exchange rate, but only one degree of freedom: once one of them is chosen, the other is endogenously determined. If the government uses money supply as a policy instrument, the exchange rate is determined endogenously (as chosen in this model). Nevertheless, the government could use the exchange rate as instrument, fix it, and then the money supply will be determined endogenously, according to the needs of cash balances of the social actors, as indicated in the second equation of system (4.13).

The government can have other instruments, depending on the model. If we introduced the bond market in the model, the interest rate would be another endogenous variable. Given that in an open economy there would be free mobility of financial capital, the international interest rate would appear as another exogenous variable. Foreign exchange will move in or move out of the economy depending, among other things, upon the differentials between domestic and international interest rates. The decrease in the international interest rate will have the same effect of an increase in the international terms of trade: the inflow of foreign exchange will rise; then both the exchange rate and the price level will tend to fall.

In this case, the government would have three possible policy instruments: the quantity of money, the nominal exchange rate, and the interest rate. However, there would still be one degree of freedom only: the government can choose one of them, and the other two will be determined endogenously.

In the epsilon model, government behavior can affect total output and distribution using monetary policy only ("helicopter money" only). However, the model is not as restrictive as it appears. In other models, the government could be using other nominal variables as instruments, but the relation between nominal variables and real variables will correspond to that predicted by the epsilon model: in the short run changes in nominal variables affect the real variables, but not to the point in which equilibrium with full employment is reached. General equilibrium with positive rate of unemployment is the outcome in the epsilon society.

A new exogenous variable appears in the general equilibrium model as a consequence of aggregation: the initial inequality (δ). Its effects are analyzed now.

Changes in the initial inequality would occur if individual endowments of economic assets are redistributed exogenously. An increase in this variable implies a higher degree of concentration of economic assets. This would occur if the capitalist class becomes smaller in size or, given its size,

the concentration of the capital stock increases within the class. Note that in this model, capitalists are endowed with human capital as well, so they receive wages and profits. Therefore, capitalists who lose capital stock ownership will become pure workers and continue receiving wage income.

The effect of an increase in the initial inequality upon output will be nil because production efficiency is independent of the degree of inequality in society (an assumption to be relaxed later on). As a result, the *functional distribution* of total output will not change either: the share of profits and wages in national income will remain unchanged; however, the *personal distribution* of income will increase because the same amount of profits will go to fewer capitalists: the average income of capitalists will increase, while the average income of workers (the real wage rate) will remain constant.

The relation between changes in the initial inequality and the degree of inequality in the personal distribution of income will be direct: the higher the inequality in the distribution of the *stock* of capital, the higher the inequality in the personal distribution of the *flow* of incomes.

The reduced form equations for total net output or national income (Y), unemployment rate (u), and the degree of inequality in its personal distribution (D) derived by the comparative statics method are as follows:

$$Y^0 = F^{\varepsilon}\ (S_m,\ z^*;\ \delta,\ K_b,\ S_h), \tag{4.16}$$

$F_1>0,\ F_2>0,\ F_4<0,\ F_5=(?),\ F_3=0$

$$u^0 = G^{\varepsilon}\ (S_m,\ z^*;\ \delta,\ K_b,\ S_h) \geq u^* > 0 \tag{4.17}$$

$G_1<0,\ G_2<0,\ G_4>0,\ G_5=(?),\ G_3=0$

$$D^0 = H^{\varepsilon}\ (S_m,\ z^*;\ \delta,\ K_b,\ S_h), \tag{4.18}$$

$H_3>0$, and $H_i = (?)$ otherwise.

Equation (4.16) indicates that variations in factor endowments determine the *level* of national income in the long run, with short-run variations around this level due to the effect of changes in the two exogenous variables money supply and international terms of trade. Equation (4.17) shows the determinants of the equilibrium unemployment rate, which must have a positive value. Finally, equation (4.18) indicates that the initial inequality in the distribution of economic assets among the population (δ) determines the *level* of inequality in the personal distribution of national income, with short-run variations around that level due to changes in the two exogenous variables, money supply and international terms of trade.

It should be noted that, from the way the epsilon model has been constructed, the macrobehavior of the epsilon economy shown above has

microfoundations and that, at the same time, the underlying microbehavior of social actors has macrofoundations. The model thus provides unity of knowledge. The partial derivatives of the system (4.16) to (4.18) constitute at the same time the causality relations, the beta propositions, and the empirical predictions of the epsilon model. They can be used to submit the theory to the falsification process.

Empirical Consistency: First World Countries

Does the abstract epsilon society resemble well the group of First World countries? The most notable empirical regularities of the First World countries in the short run were listed in chapter 2. The beta propositions of the epsilon static model presented here can now be confronted against the short-run facts, which include Facts 1 and 4.

The epsilon model predicts that any equilibrium situation will imply a positive rate of unemployment ($u>0$) because unemployment plays a role in the economic process: capitalism needs unemployment to operate. This equilibrium condition is observable and, thus, it makes the model refutable. The epsilon model indeed predicts the existence and persistence of unemployment (Fact 1). The epsilon model also predicts the interplay between nominal and real variables in the short run (Fact 4). According to the epsilon model, changes in the money supply (or in the nominal exchange or in the nominal interest rate) will affect output and employment.

The predictions of the epsilon model are also consistent with some of the empirical results of standard macroeconomics. In this literature, the modified Phillips curve shows an inverse relation between changes in the inflation rate and the unemployment rate, which in the case of the US economy, for the period 1976–2008, shows that the *minimum* unemployment rate observed is around 4 percent. If the inflation rate was utilized instead of the price level, the epsilon model would also predict the Phillips curve together with nonzero minimum unemployment rate. Moreover, the *average* unemployment rate for a long period (1970–2008), varies across countries: 2.3 percent in Japan and 6.1 percent in the United States (Blanchard 2009, p. 190). The epsilon model would interpret these figures by the differences in the necessary unemployment rates (u^*) between these countries: labor relations are less conflictive in Japan than in the US economy.

However, contrary to standard macroeconomics, the epsilon model explains why unemployment exists and persists in the First World: unemployment in the labor market is a necessity for the functioning of the system. On the other hand, labor markets do not operate with subsistence real

wages, as the classical theory assumed. If real wages covered the needs of workers, poverty would hardly exist in the First World. According to the epsilon model, workers do not earn what they need for living; rather they seek to make a living with what they earn. Limits to the real wage rate are determined by efficiency wages, not by subsistence wages.

In sum, the two fundamental facts of the First World countries (Facts 1 and 4) do not refute this model of the epsilon theory; hence, there is no reason to reject epsilon theory and we may accept it provisionally at this stage of our research. Partial theories that intend to explain production and distribution in the Third World will be presented in the next two chapters.

CHAPTER 5

The Omega Society

In Third World countries, workers who are out of wage employment are mostly engaged in self-employment. They produce goods and services in small units of production that are located in rural areas (in small farms), and in urban areas (in small shops, some of them on the streets). The income they make in these small units is, in general, lower than the average wage rate being paid in the labor market for similar skills, that is, self-employment implies underemployment. This is Fact 2 listed in chapter 2.

How can one explain production and distribution in Third World countries? Surely, by constructing an abstract society, an economic theory. An abstract capitalist society—called omega—will now be constructed and then submitted to empirical refutation.

Omega: Socially Homogeneous Class Society and Overpopulated

Omega society will also constitute a class society. Its social institutions are thus similar to those of the epsilon society. The only difference rests upon factor endowments: omega is an *overpopulated* society. Analytically, overpopulation means that the marginal productivity of the total labor force is near zero.

Could omega society operate as an epsilon society? As shown in the previous chapter, epsilon operates necessarily with unemployment; hence, the only effect of an increase in the labor supply would be the increase in unemployment. If the labor supply continued to increase, unemployment would be so large that the labor market would not be able to function with the prevailing institutional norms. If the government were paying unemployment

insurance, the financing of this program would become unviable. The excess labor supply would be well beyond the necessary rate of unemployment (u^*); so unemployment would be too large for the unemployed workers to make a living just by searching for jobs and finding jobs.

In sum, in an overpopulated economy, unemployment cannot operate as the device to discipline workers. Therefore, omega will function differently from epsilon. To show how the omega society will function is the objective of this chapter.

The alpha propositions for the omega society (ω) include all the propositions that were established for the epsilon economy, except that, as initial conditions, there will exist overpopulation in the labor market. For easy reference the set of alpha propositions is fully listed as follows:

- $\alpha(\omega)$. *(1) Institutional context*: (a) Rules: People participating in the economic process are endowed with economic and political assets. Economic assets are subject to private property rights. Individuals can exchange goods subject to the social norms of market exchange, which include the norm that nominal wages in labor markets cannot fall; the market system operates with Walrasian and non-Walrasian markets, in which the labor market is of the latter type. The political regime is democratic. (b) Organizations: households, firms, and the government.
- $\alpha(\omega)$. *(2) Initial conditions*: (a) Individuals are endowed with unequal quantities of economic assets but with equal political entitlements. There are two social classes: capitalists and workers. (b) Factor endowments: the stock of capital per worker is such that the marginal productivity of the entire labor force is near zero, that is, there is overpopulation.
- $\alpha(\omega)$. *(3) Economic rationality of agents*: Consistent with the institutional context of capitalism, individuals act motivated by self-interest. Capitalists seek two particular objectives, hierarchically ordered: first, maintenance of class position and, second, maximization of profits. In the labor market, workers seek to maximize wages and minimize effort, while capitalists seek to minimize wages and maximize effort. Due to this conflict in labor relations, capitalists use underemployment as the device to extract effort from workers. Individuals have a limited tolerance for inequality.

A Static Model of the Omega Theory: The Short Run

In order to make omega theory falsifiable, an omega model must be constructed, which can be seen as constructing a particular stage set in which

social actors play their roles. The following auxiliary assumptions are introduced for that purpose:

- There are two social groups: capitalists who are endowed with stocks of physical capital and human capital, and workers who are endowed with stocks of human capital and cash balances. Wages are paid at the end of the production period; consequently, workers need to hold cash balances for transaction purposes. Government behavior consists of supplying money to the economic process.
- There are two production sectors: the capitalist sector, producing with hired workers, and the subsistence sector, in which workers are self-employed. The discipline device in the labor market is based on the gap between the wage rate and the marginal income in the subsistence sector.
- Four markets constitute the market system: labor, commodity, money, and foreign exchange. The labor market is non-Walrasian, but the rest are Walrasian. Human capital is the same for all workers, so there is only one labor market. One sole good (B) is produced in society. Workers and capitalist firms seek to exchange labor services and goods in the labor and commodity markets. Capitalist firms seek to exchange goods in international markets, using foreign exchange as the means of payment. There is no market for renting capital services.
- Capitalists, workers, and the government interact and compete in the market to obtain their objectives, but none has the power to set prices (a perfect competitive market system); they all face costless information on market prices and technology. Labor productivity levels vary across firms.
- The economic process is static and open.

The economic structure of the omega economy consists of a capitalist sector and a subsistence sector; hence, there are two types of production units, capitalist firms and subsistence units in which workers are self-employed. Production technology in the capitalist sector uses capital and labor, whereas in the subsistence sector it uses only labor. The labor productivity level of the subsistence sector is, therefore, lower than that of the capitalist sector; moreover, productivity levels are too low here to allow the units to operate as a capitalist firm (to pay wages and generate profits), and it is called subsistence sector in this sense. In a world of one good, output in the subsistence sector will be destined to direct consumption by the producers, not to market exchange.

The Lewis Model: Horizontal Labor Supply Curve

Arthur Lewis (1954) developed a model for an overpopulated society, which is going to be presented here as the first model of the omega theory. He introduced the assumption that, in order to extract effort from workers, firms will seek to pay a wage rate that is above the wage earners' opportunity cost, which is equal to the income that they can make in the subsistence sector. This gap will measure the cost wage earners will pay if they are dismissed from the firm due to shirking behavior. The value of the premium is exogenously given.

The Lewis model also assumed that the average productivity of labor in the subsistence sector is constant, which implies that average productivity and marginal productivity of labor are equal. The real wage rate is equal to a premium (say 30 percent) above the constant average productivity of labor in the subsistence sector; therefore, the real wage rate in the labor market is exogenously determined. If peasants constitute the subsistence sector, the model will predict that the average productivity of the peasant economy determines the real wage rate in the urban labor market. Hence, the real wage rate will rise if and only if the labor productivity level in the peasant economy increases.

Given the real wage rate, capitalist firms will seek to hire workers in a quantity that makes profits the largest. The quantity of employment in the labor market is determined by demand alone. Due to overpopulation, at this real wage rate, there will be excess supply of labor, which will be self-employed in the subsistence sector. Any size of excess labor supply will become self-employed in the subsistence sector for the capacity of absorption of this sector is unlimited. This is how the labor market operates in this model.

However, the empirical regularities listed in chapter 2 refute the empirical predictions of the Lewis model. First, this model predicts zero unemployment in the short run. This prediction is inconsistent with the reality of the Third World countries, Fact 2, where we observe unemployment, in addition to a significant amount of self-employment. Second, in the long run, the model predicts that real wages cannot rise if capital stock increases. This prediction is inconsistent with Fact 5. Third, real wage and employment are both determined in the labor market by real variables alone, such as capital stock, technology, and labor productivity in the subsistence sector. Nominal variables, such as money supply, do not play any role. If incorporated into a general equilibrium system, this labor market model would predict money neutrality, which is inconsistent with Fact 4: nominal variables are correlated with real variables.

A New Model: Rising Labor Supply Curve

The new model will abandon the auxiliary assumption of constant average productivity of labor in the subsistence sector and will assume diminishing returns instead. Production units in the subsistence sector have different labor productivity levels due to differentiated access to location. In this case diminishing returns arise not because of the problem of overcrowding a fixed production factor, such as overcrowding with labor the existing capital stock (there is no capital here), but because of quality differences between production units. Additional units entering into the subsistence sector will have to work in locations of lower quality, such as longer distance to public goods (infrastructure).

The model thus assumes *Ricardian diminishing returns* in the subsistence sector, which is the result of using locations of lower quality as more units enter into the subsistence sector. Figure 4.1 panel (b), chapter 4, can represent this property of Ricardian diminishing returns in the subsistence sector if the curve now refers to subsistence units, instead of capitalist firms; and the area under the marginal productivity curve (total income) is equal to labor income, instead of profits and wages.

The implication of this assumption is that marginal productivity of labor in the subsistence sector will be diminishing. Additional self-employed workers will add smaller and smaller quantities to total output in the subsistence sector; consequently, average productivity of labor will also decline as more workers that are self-employed enter into the subsistence sector.

Workers face the alternative of taking the jobs in the capitalist sector or becoming self-employed in the subsistence sector. Assuming that workers are homogeneous and have no special preferences for either alternative, the equilibrium allocation of total labor into the two sectors would be when the marginal productivity of labor in the subsistence sector equals the market real wage rate. So for a given wage rate, there will be a number of workers willing to take the job in the capitalist sector, namely, those unable to make a higher income in the subsistence sector. If the real wage rate were higher, then the group of workers unable to make higher income in the self-employment sector would increase, which implies an increase in the number of workers seeking wage employment. Thus, we have just derived the labor supply curve: The higher the wage rate, the higher the quantity of labor supplied to the labor market. Therefore, the marginal productivity curve of the subsistence sector represents, at the same time, the labor supply curve.

Capitalist firms will seek to pay efficiency wages: a wage rate that is higher than the opportunity cost of labor, which is given by what the income workers can make in the subsistence sector, that is, the marginal productivity of

labor in the subsistence sector, which in turn is the labor supply curve. This gap is a premium that intends to be the labor discipline device. The wage premium will be applied to the rising labor supply curve. When shifted upward by the proportion of the premium, the labor supply curve will then become the *effort extraction curve*. This curve will operate as a restriction for wage determination in the labor market, that is, the minimum value of the efficiency wage set (w^*) will be determined at the crossing of the labor demand curve and the effort extraction curve.

It follows that w^* will be higher than the Walrasian wage rate (where demand and supply curves meet). It also follows that the solution of price and quantity of the labor market is not independent of the marginal productivity of labor in the subsistence sector, that is, an exogenous change in this curve will modify the supply curve of labor and the effort extraction curve.

The selfish motivations of capitalists and workers that were assumed in the epsilon society will also be applied to the omega society. However, those motivations will be pursued under different constraints.

Behavior of Capitalists

Capitalists seek profit maximization but in the particular social context of omega. For the typical firm, this motivation and the constraints can be written as follows:

$$\text{Maximize } P = P_b\, Q_b - P_c\, D_c - P_h\, D_h \tag{5.1}$$

subject to

$Q_b = \lambda\, \phi\, (D_h, K_b)$, where $\lambda = \lambda^* = 1$
$D_c = c\, Q_b$
$X_b = (1/z^*)\, D_c$
$w \geq w^* > w^w$.

This is the same set of structural equations that were established in the epsilon model—system (4.7), chapter 4.

The equilibrium condition of the individual firm is therefore similar to the condition that was obtained in the epsilon model, namely, the market nominal wage must be equal to the value of the net marginal productivity of labor in the firm, where "net" means total output minus the output utilized to pay for the required imported inputs, the good C. This equilibrium is (trivially) stable. The comparative statics at the individual firm level and

then the further aggregation over all firms will be similar to the procedure shown in equations (4.9) and (4.11). The latter is reproduced now as the following system:

$$D_h = H\,(P_h,\ P_c,\ P_h,\ K_h),\ H_1>0,\ H_2<0,\ H_3<0,\ H_4>0 \tag{5.2}$$

subject to

$$P_h\,S_h = (P_c\,D_c - P_h\,X_h) + P_h\,D_h$$
$$X_h = (1/z^*)\,D_c$$

subject also to the following social constraints

$$P_h \geq P_h^*,$$
$$w \geq w^* = [(1+p)\ \mathrm{v}'] > w^w.$$

In system (5.2), the first equation is the aggregate labor demand function. The first group of constraints includes the aggregate budget constraint of firms and the aggregate export function of firms. The second group of constraints includes the social norm that the nominal wage cannot fall and the condition related to ensure labor discipline: The market real wage rate must be equal to or higher than the minimum efficiency wage (w^*), which in turn must be equal to the marginal productivity of labor in the subsistence sector plus the premium ($1 + p$); moreover, this minimum efficiency wage must be higher than the Walrasian real wage rate (w^w).

Behavior of Workers

Workers are endowed with human capital and cash balances. They seek to maximize total real income subject to their resource endowments and the rationing mechanisms of the labor market.

As a first option, workers seek employment in the capitalist sector, for the market wage rate is higher than the income they can make in the subsistence sector. The incentive system leads them to this endeavor, but jobs for all workers are not available, no matter how hard each worker seeks employment in the capitalist sector. Because there is a turnover of workers in the capitalist sector due to dismissals of those workers found shirking, there is a probability of finding a job. Assume this probability is the same for all workers.

Let w^e represent the workers' expected wage if seeking a job. For an exogenously given probability π of finding a job, the expected wage depends

positively on the market wage rate. These assumptions imply a uniform expected wage for all workers, which can be written as

$$w^e = \pi\, w, \text{ where } 0 < \pi < 1 \tag{5.3}$$

As a second best solution, the workers who become excluded from employment in the capitalist sector will choose between unemployment and self-employment. These workers will thus evaluate the expected wage (what they would get after seeking and finding a job) against the sure income they can make in the subsistence sector.

For a given market real wage rate, therefore, the redundant workers will face an expected income if they chose to be unemployed. On the other hand, the higher the number of workers that choose self-employment, the lower the income they can make, due to diminishing returns; therefore, for a number of these workers the expected wage rate would be lower than the income they can make in the subsistence sector, but for others it would be higher. Assuming that redundant workers are homogeneous and have no preferences in favor of any of the two residual options, the equilibrium allocation of workers between unemployment and self-employment will take place when the expected wage rate is equal to the marginal productivity of labor in the subsistence sector.

Hence, a group of the excluded workers would choose to seek jobs in the labor market, whereas the rest would choose self-employment in the subsistence sector. The group who chooses to seek a job becomes *voluntarily* unemployed, as self-employment is an open alternative. However, note that this is a choice under second best alternatives. The first option, the workers' most preferred situation, is to have a job in the capitalist sector. The model assumes transitions of workers between employment in the capitalist sector, unemployment, and self-employment in the subsistence sector, whenever jobs are increasing or decreasing in the capitalist sector.

Because workers prefer wage employment than the other two alternatives, the behavior of *wage earners* will be similar to the behavior presented in the epsilon model. This is reproduced here just for convenience as

$$D_b = F(P_b, D_h, P_h),\ F_1<0,\ F_2>0,\ F_3>0 \tag{5.4}$$
$$D_m = P_b\, L(D_b),\ L'>0,\ L''<0$$

subject to the aggregate budget constraint of wage earners

$$P_b\, D_h = P_b\, D_b + (D_m - S_m).$$

General Equilibrium

The Capitalist Sector

There are four markets that operate under perfect competition in the capitalist sector. The conditions of equilibrium for each market are

$$\text{Labor market } S_h \equiv D_h + E_h,\ P_h \geq P_h^*,\ w \geq w^* > w^w \tag{5.5}$$
$$\text{Commodity market } S_b = D_b$$
$$\text{Foreign exchange market } P_c^*\ D_c = P_b^*\ X_b$$
$$\text{Money market } S_m = D_m$$

subject to the aggregate budget constraint

$$P_b\ (D_b - S_b) + P_e\ (P_c^*\ D_c - P_b^*\ X_b) + (D_m - S_m) = 0.$$

The labor market is non-Walrasian; it operates with excess labor supply ($E_h > 0$) and is subject to the constraints that the nominal wage cannot fall and the real wage must lie within the range of efficiency wages. The rest are Walrasian markets.

The market system will solve for the values of the endogenous variables, for given values of the exogenous variables. These variables are

Endogenous variables: P_e, D_h, P_b , P_h, w, E_h, Q_b, Y, W, P

Exogenous variables: z^*, r^*, P_h^*, K_b, S_h, δ, S_m.

The set of endogenous variables includes some variables that were exogenous in the microlevel of analysis or partial level of analysis, such as nominal prices. This is in accord with the principle of increasing endogenization of variables in the aggregation process.

In order to determine the core of the general equilibrium, we proceed as in the epsilon model. Equilibrium in the foreign exchange market is incorporated into the labor market equation. That leaves us with two Walrasian markets (good B and money), of which one can be eliminated by invoking Walras's law. Then general equilibrium can be established by solving two markets, say the money market and the labor market. Then the core of the system is given by the following equations

$$D_h = \mathrm{H}\ (P_e; P_h, z^*, K_b),\ H_1 > 0,\ H_2 < 0,\ H_3 > 0,\ H_4 > 0 \tag{5.6}$$
$$S_m = P_e\ P_b^*\ L\ (W) = M\ (P_e, D_h; P_h,\ P_b^*),\ \text{where } M_i > 0, \text{ for all } i$$
$$\text{subject to } w \geq w^* = [(1+p)\ v'] > w^w.$$

The core is constituted by two equations and two endogenous variables, the exchange rate P_e and the employment level D_h. The solution is simultaneous. Once these values are known, the rest of the endogenous variables can be solved just by implication.

Figure 5.1 depicts the core of the general equilibrium. Panel (a) shows the labor market, in which the demand curve h is presented in relation to the nominal wage rate. The position of this curve is determined once the exchange rate is known. Panel (b) shows the money market, in which the particular position of the demand curve M is determined once the wage employment level is known. Hence, there is a pair of values that determine the equilibrium values of these endogenous variables. The solution is simultaneous,

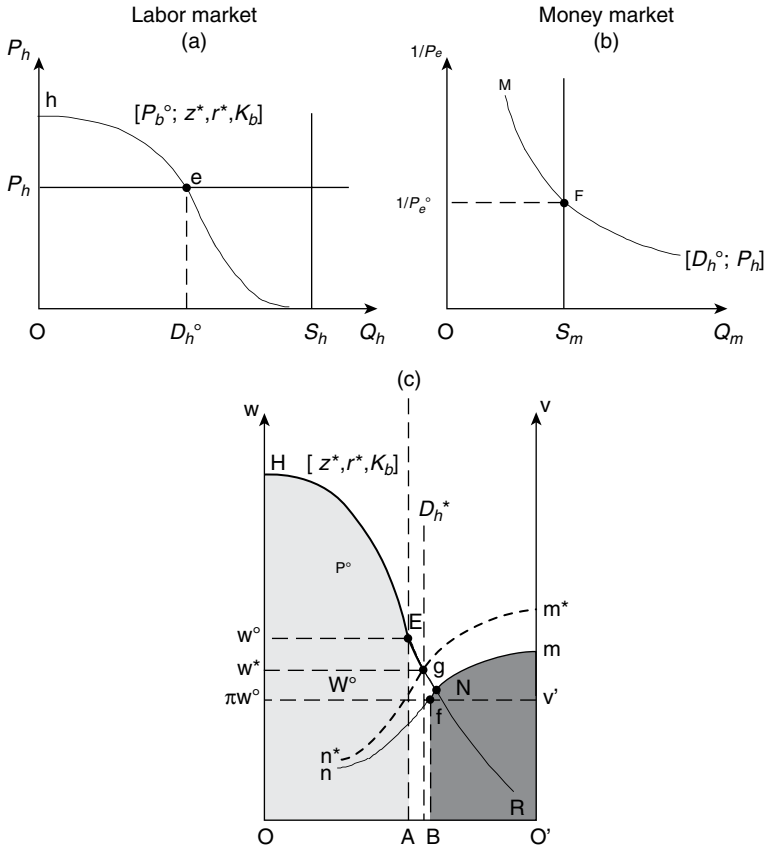


Figure 5.1 General equilibrium in omega society.

marked by point e and point F. Once the exchange rate is known, the price level is determined, so is the real wage rate of equilibrium.

Figure 5.1 panel (c) presents the equilibrium of the labor market in terms of real wage rates. The curve nm represents the labor supply curve, which is transformed into the effort extraction curve n*m*, just by adding the value of the premium needed for labor discipline. As indicated by the efficiency wages constraints of the system (5.6), only the segment Hg of the curve HR represents the labor demand curve. The market real wage rate must belong to the set of efficiency wages; thus there exists a real wage w* that is the minimum value that the real wage can take in order to maintain the labor demand curve fixed (the level of the marginal productivity of labor). Graphically, the point at which the effort extraction curve and the labor demand curve cross each other, point g, determines this minimum wage. The value of w^* is the lowest of the set of efficiency wages and determines the maximum wage employment level D_h^*.

If for some reason, the market real wage rate took the value of full employment, point N, the curve HR would shift downward; hence, point N is unattainable. Therefore, point E indicates an equilibrium position (the counterpart of point e in panel (a)). The equilibrium employment level is OA and the equilibrium real wage is w^o. The excess labor supply is equal to AO′, where OO′ is the total labor supply.

Among the Walrasian markets of the system, the money market is explicitly in equilibrium. The foreign exchange market is implicitly in equilibrium because it is incorporated into the labor demand curve. It can be shown that the commodity market is also in equilibrium. The quantity of good B supplied to the domestic market is equal to the difference between total net output (net of exports to pay for the technological required inputs) and profits; in figure 5.1 panel (c), this is equal to the area of the wage bill, which in turn is equal to the quantity demanded. Thus, in fact, there exists general equilibrium in the market system.

The Subsistence Sector

Once the values of equilibrium in the labor market have been determined, the quantity of the excess labor supply will be known. Then the allocation of the redundant labor into unemployment and self-employment needs to be solved. The relevant equations to solve this allocation include

$$E_h^0 = S_h - D_h^0 = L_s + U \tag{5.7}$$

$V = J(L_s)$, where $J' > 0$, $J'' < 0$

$\pi w^0 = J'(L_s) \equiv v'$, where $0 < \pi < 1$ and $\pi w^0 > J'(E_h^0)$.

The first equation points out that in the omega economy, excess labor supply takes the form of unemployment and self-employment. (This is different from the epsilon economy, where excess labor supply takes the form of unemployment only.) The second equation is the production function in the subsistence sector, which is subject to Ricardian diminishing returns. These two are the structural equations. The third equation shows the equilibrium condition for the allocation of labor to self-employment: the marginal productivity of labor in the subsistence sector (v') must be equal to the expected wage rate (πw^0).

Given the marginal productivity of the labor curve in the subsistence sector, the self-employment level will be determined by the expected wage rate; given the total amount of excess labor supply, the unemployment level of equilibrium will be determined once the self-employment level is determined. Thus, the unemployment level is a residual of the residual labor. But unemployment could even be zero in equilibrium, which will occur when the expected wage rate is equal to or lower than the marginal productivity of total excess labor supply and thus there is no incentive to become unemployed; alternatively, we can say that unemployment will be positive if the expected wage rate is higher than the marginal productivity of total excess labor supply, which is the condition that appears in the third equation of system (5.7)

The static general equilibrium solution of the omega model is clearly sequential. First, equilibrium in the capitalist sector is determined, which determines the excess labor supply as well; then, equilibrium self-employment in the subsistence sector is determined, which in turn determines the amount of unemployment.

Figure 5.1 shows the general equilibrium of the omega economy. Once the capitalist sector solution is attained by the interactions of the labor market and the money market as shown in panels (a) and (b), wage employment level and the real wage rate will be known, as shown in panel (c). Then the expected wage is determined. This is marked as a fraction of the market wage rate. This value must be equal to the marginal productivity of labor, which occurs at point f; hence, the self-employment level is determined. Unemployment is the residual of the total excess labor supply. Thus, the allocation of the excess labor (A0′) to unemployment (AB) and self-employment (B0′) is solved. In order to assure labor discipline and keep the curve HR unchanged, the gap between the real wage and the marginal productivity of the self-employed workers must be equal to or higher than the premium rate, as at point E. General equilibrium is thus achieved.

The labor market is clearly a non-Walrasian market. Workers are willing to exchange their labor at the market real wage, but not all are able to do so. Unemployment or self-employment is the result of economic choice made by those workers excluded from the labor market. In this model, as pointed out

above, workers who choose to seek a job in the capitalist sector become voluntarily unemployed in the sense that it is the result of economic choice, for they take the lower self-employment income. But it is involuntary in the sense that unemployment is not the most preferred situation for workers. Unemployment or self-employment is an alternative under a second-best situation. Because self-employment generates income that is below the market wage rate for the same skill endowments, self-employment is called *underemployment*.

As can be seen in figure 5.1 panel (c), the equilibrium rate of unemployment need not be a positive number; it could be zero. In omega, a labor-abundant society, unemployment is not a necessity for the functioning of the capitalist system; the variable of underemployment plays that role. Underemployment is necessary for labor discipline and thus for high labor productivity and high profits in the capitalist sector. Workers that are not needed in the capitalist sector get self-employed to generate their *own* income, which makes viable the capitalist sector in an overpopulated society. Underemployment, not unemployment, is a necessity for the functioning of capitalism in an overpopulated society.

As shown in chapter 3, Professor Robert Lucas (1978), referring to the First World (epsilon society), argued that unemployment is mostly voluntary: workers decide to seek for a job instead of taking up any other job or occupation; hence, unemployment does not imply excess labor supply. But all the same, Lucas assumes Walrasian labor markets, which makes his proposition logically inconsistent, for there would not exist any labor market in which the unemployed could take up any job. In light of the omega model presented here, however, Lucas's argument could apply to an overpopulated society in which the workers who are excluded from the labor market "at any time can always find some occupation at once" (1978, p. 354), but as self-employed in the subsistence sector.

To be sure, in the omega model, unemployment is voluntary in the sense that it is the result of an individual choice (between unemployment and self-employment), but it is involuntary in the sense that it is a second best choice. In the omega model, the excess labor supply takes the form of both unemployment and self-employment (or underemployment).

Production and Distribution

National income (Y) of equilibrium and its distribution between social groups can be represented as follows:

$$Y = P + W + V = P + w\,D_h + v\,L_s \tag{5.8}$$
$$S_h = D_h + L_s + U.$$

The term v measures the average productivity of labor in the subsistence sector.

Figure 5.1, panel (c), shows the national income of equilibrium and its distribution. National income is equal to the aggregation of two areas: the area under the curve HE (total net output in the capitalist sector) plus the area under the curve mf (total output in the subsistence sector). Capitalist firms and the small units in the subsistence sector produce the same good. In this world of one commodity, this aggregation is simple and fully justified.

The static general equilibrium solution is clearly inefficient. The allocation of labor given by the intersection of the supply and demand curves would maximize national income. This equilibrium occurs at point N in figure 5.1 panel (c), where the labor supply curve is given by the curve nm, measured from origin 0. However, this solution is not economically attainable. It would require a Walrasian labor market. The omega economy is inherently inefficient. In comparison, it was shown in the previous chapter that the epsilon economy is also inherently inefficient due to the necessary existence of unemployment.

As to the equilibrium income distribution, the area under the curve HE is the total net output generated in the capitalist sector, which is distributed between profits and wages. The area under the curve mf is the total output generated in the subsistence sector. The functional income distribution is thus determined. To determine the personal income distribution, we must take into account the hierarchy of incomes among workers. The market wage rate is higher than the average income of the self-employed, which would lie above the value of the marginal productivity of O′B workers, that is, above point f, but below point m, which is lower than the market real wage rate. The unemployed has an expected wage that is equal to the marginal income in the subsistence sector.

To be sure, the model assumes that units of production where the marginal productivity of labor can be equal to or higher than the real wage rate will operate as capitalist firms because profits (the gap between average productivity and marginal productivity) will be feasible. This explains why the marginal productivity curve of the subsistence sector starts from the point m, which is below the market real wage rate. This also explains the name "subsistence sector," which means that profits and hired labor will be unfeasible.

In this model, all workers are homogenous, endowed with the same human capital. Yet, equilibrium occurs with inequality among workers. This result may seem paradoxical. According to the model, those workers who are excluded from the labor market make up the poorest groups of society. Can they not offer their labor at lower nominal wages to get wage

employment? No, they cannot. Social norms impede this type of competition among workers. Furthermore, in an overpopulated society, this competition would lead to equilibrium with the Walrasian wage rate, which is near zero. This is socially unviable!

In addition, the motivation of capitalists is not employment maximization, but profit maximization, which leads to a solution with a floor on wage rates and a maximum amount of employment, which is not full employment, with uniform wage rates. The inequality among workers originates in the labor discipline device that is needed to maximize profits. Capitalists acting under the motivation of self-interest lead to the aggregate result of inequality among workers.

Workers who are excluded from the labor market can also seek to set up firms or expand the size of their current productive units where they are self-employed. Skill level does not present a barrier. They could rent capital to open firms, but there is no market for renting capital service. They would have to buy physical capital, which in turn needs financing; however, they are also excluded from the credit market and the insurance market, making these projects unviable.

Credit and insurance markets operate under asymmetric information conditions, which imply significant transaction costs and complex market equilibrium conditions (Stiglitz 2002). The existence of economies of scale in the loan size explains the fact that credit markets exclude the poor and the small business. Therefore, banks will have incentives to deal with capitalist firms only, where the loan sizes are larger; then, only those eligible to participate in the market together with banks will determine prices and quantities; hence, the bank credit market operates as quasi-Walrasian (Figueroa 2011). The same principle of economies of scale can also explain the exclusion of small businesses from the insurance market: Transaction costs require minimum insurance size for the market exchange to be profitable.

Hence, to set up new firms, and become capitalists, is not an option that workers may use as a mechanism to escape equilibrium with inequality, including inequality among workers, in the omega model. There is no mechanism by which workers can become capitalists if they want to. Therefore, inequality among homogeneous workers is an equilibrium situation in the omega model, for no social actor has both power and incentive to change it.

The static general equilibrium implies unemployment and underemployment, together with inequality, including inequality among homogenous workers. In the static system, the values of general equilibrium will be repeated period after period, as long as the exogenous variables of the system remain unchanged.

Beta Propositions

The sequential solution of the general equilibrium in omega society implies that the empirical predictions of the model can independently be analyzed for the capitalist sector. If the market equilibrium that has been attained in the capitalist sector is stable, comparative statics can be applied to derive beta propositions. The core of the market equilibrium is indeed stable, for the money market is stable and the labor market is non-Walrasian. Comparative statics can thus be applied to this equilibrium in order to derive beta propositions: the effects of exogenous variables upon endogenous variables.

The endogenous and exogenous variables in the capitalist sector of the omega model are the same as those of the epsilon model, and so are the structural equations; therefore, the effects of exogenous variables upon endogenous variables in the capitalist sector of the omega model would be similar to those presented for the case of the epsilon model. The relevant exogenous variables in the short-run analysis include money supply and terms of trade.

First, consider an increase in the money supply. This will generate directly an excess of cash balances willingly held by workers, who will then intend to restore the equilibrium by using this excess to buy more quantities of good B in the period of adjustment. This adjustment would reduce both the exports and foreign exchange inflows and thus an increase in the exchange rate. Then the price level would increase and the real wage would fall; hence, firms would seek to hire more workers (from the pool of excess supply) and produce more output to obtain more profits. This would be the first round effect. The second round effect will be ambiguous because the fall in the real wage rate and the increase of employment imply an ambiguous change in the real wage bill, which in turn implies an ambiguous and small change in the demand for real cash balances and on the price level. Therefore, the second round effect upon the price level would be small and may be neglected. The effect of an increase in money supply is to increase employment by decreasing the real wage rate.

The effect of an expansionary monetary policy on output can be visualized with the help of figure 5.1 panel (c). Point E represents the initial general equilibrium. An increase in the money supply increases the price level, which implies a fall in the real wage rate and an increase in the employment level. The initial equilibrium situation, at point E, will be moved to another equilibrium situation, which will lie below point E, in the segment Eg. The government can then reduce the excess supply of labor using monetary policy. The lower real wage rate leads to lower expected wage rate, whereas the higher employment level leads to lower excess labor supply and higher labor

marginal productivity; thus, incentive for becoming unemployed will tend to fall; it may even become zero.

A sufficient large increase in money supply will ultimately reach the effective full employment level, at point g, where we assume unemployment is zero. If money supply keeps increasing, employment will remain fixed, and only the exchange rate and consequently the price level will rise, which will then induce increases in the nominal wage rate in the same proportion in order to avoid the fall in real wage below the minimum efficiency real wage rate (w^*), that is, *money becomes neutral*. It should be clear that the Walrasian market equilibrium, point N, is unattainable in the omega model.

Changes in income inequality are determined by changes in the concentration of total output by the richest groups of society. The usual measure is the Gini coefficient, which is based on the Lorenz curve. Income inequality increases when the Lorenz curve shifts outward, which implies that the income share of the richest social groups increases or when the income share remains constant, but the relative size of these social groups is smaller; alternatively, this occurs when the relative average income of the richest groups increases or when the relative income remains constant, but the relative size of these social groups is smaller.

On income distribution, the effect of an increase in money supply is as follows. With respect to the initial general equilibrium, at point E and point f, the new production equilibrium shown implies higher total output both in the capitalist sector and in the subsistence sector. However, the change in the wage bill is ambiguous. Hence, change in the share of wages and that of profits in total output is ambiguous; output in the subsistence sector also increases, but its share in national income is unclear. Change in functional income distribution is thus ambiguous. As to changes in personal distribution or relative incomes, average profit of capitalists increases in absolute and relative terms, while real wage rate falls, and average real income in the subsistence sector falls as well. Nevertheless, at the same time, unemployment (with zero income) falls. Hence, Lorenz curves will tend to cross each other and overall income inequality will tend to remain unchanged, or changes will be small, that is, the Gini coefficient will tend to remain unchanged or the change will be small.

Now consider the effects of an increase in the international terms of trade. As can be visualized in figure 5.1 panel (c), an increase in the international terms of trade (z^*) shifts the marginal labor productivity curve HR outward. The initial general equilibrium is at point E. At the current real wage rate, firms would seek to hire more workers (drawn from the pool of excess supply); hence, the market real wage rate could remain unchanged. However, this is not the case. The wage bill would increase. Given the quantity of

money supplied, the buildup of cash balances would imply a temporary fall in the quantity of good B demanded, which would increase both exports and the inflow of foreign exchange leading to a fall in the exchange rate, which in turn would lead to a fall in the domestic price level and a rise in the real wage rate; hence employment would still increase. Then the initial equilibrium situation, given by point E, will be moved to another equilibrium situation, say point E″ (not marked), which will be located to the northeast of point E, where both employment and real wage rate will be higher.

What is the subsequent effect of changes in the international terms of trade upon the subsistence sector? The effect is indirect, through the effect upon the labor market. The increase in the international terms of trade implies an increase in both the real wage rate and wage employment, which in turn implies a reduction in the excess labor supply. However, the increase in the real wage rate leads to higher expected wage rate and thus to a decrease in self-employment, and thus to a fall in total output in the subsistence sector. Change is unemployment is ambiguous because the effect of higher wage employment, maintaining the real wage rate constant, is to reduce unemployment; but the increase in real wage rate has the effect of increasing it, as self-employment falls.

Figure 5.1 panel (c) allows us to visualize the effect of an increase in the international terms of trade upon income inequality. Total output increases in the capitalist sector, but it declines in the subsistence sector. Within the capitalist sector, the wage bill increases, but the change in profits is ambiguous due to the increase in the real wage rate; hence, the share of profits in national income is ambiguous, so is the share of wages. As to changes in relative incomes, the change in the average profits is ambiguous; the real wage rate and the average income from self-employment both increase, but changes in the relative incomes are ambiguous. Therefore, change in income inequality is undetermined.

The final exogenous variable is the initial inequality. Changes in the initial inequality would occur if the individual endowment of economic assets were redistributed exogenously. An increase in this variable implies a higher degree of concentration of economic assets. This would occur if the capitalist class becomes smaller in size or, given its size, the concentration increases within the class. The effect of an increase in the initial inequality on output will be nil because the production efficiency of the economy is independent of the degree of income inequality (an assumption to be relaxed later on).

With regard to income inequality, the functional distribution will remain unchanged, but personal distribution will become more unequal because the same amount of profits will become concentrated in fewer individuals. The effect of an increase in the initial inequality on the degree of income

inequality will be positive: The higher the inequality in the distribution of the *stock* of resources, the higher the inequality in the distribution of the *flow* of incomes.

Now recall that general equilibrium in the omega model includes as one of its structural equations the marginal productivity of labor in the subsistence sector, that is, *given* this curve, general equilibrium is attained. If this curve shifted upward exogenously, then the effort extraction curve would be shifted upward, and the value of w^* would increase. The solution of price and quantities in the capitalist sector would be different. This is to indicate that the labor productivity in the subsistence sector plays a role in the general equilibrium solution, although not the strong role that the Lewis model predicted.

Why is it that the real wage rate can increase in an overpopulated society? To recall, the Lewis model predicted fixed real wages, that is, an unlimited supply of labor at a given real wage rate. The real wage rate could increase only after the overpopulation had been eliminated. However, the omega model predicts that the real wage rate can increase before overpopulation has been eliminated. This apparent paradox can now be explained. If the labor market were Walrasian (as in the Lewis model), but the supply curve were rising, then increases in labor demand would also imply an increase in the real wage rate. In the omega model, as in the epsilon model, the assumptions about the labor market are that it is non-Walrasian and that the real wage rate and the employment level are not determined by real variables alone (factor endowments, physical labor productivity, technology, real opportunity cost) but also by nominal variables as well. The general equilibrium shown in figure 5.1 (c) summarizes these assumptions about the labor market. Recall that the real wage rate of equilibrium is determined once the price level is determined, for the nominal wage rate is exogenously given.

In sum, from the omega model we have been able to derive the reduced form equations for the level of national income (Y), employment in the capitalist sector (D_h), and the degree of inequality in personal distribution of national income (D), which are:

$$Y^0 = F^\omega (S_m, z^*; \delta, K_b, S_h) \tag{5.9}$$

$$F_1>0,\ F_2>0,\ F_4<0,\ F_5>0,\ F_3=0$$

$$D_h^0 = G^\omega (S_m, z^*; \delta, K_b, S_h) \leq D_h^*<S_h \tag{5.10}$$

$$G_1>0,\ G_2>0,\ G_4<0,\ G_5>0,\ G_3=0$$

$$D^0 = H^\omega (S_m, z^*; {}^\delta, K_b, S_h) \tag{5.11}$$

$$H_3>0, \text{ and } H_i = (?) \text{ otherwise.}$$

Equation (5.9) indicates that variations in factor endowments determine the *level* of national income in the long run, with variations in the short run around this level due to the effect of changes in the two exogenous variables money supply and international terms of trade. Equation (5.10) shows the determinants of the equilibrium quantity of labor demanded in the capitalist sector, which cannot take values equal to the quantity of labor supplied; hence, equilibrium in the labor market is necessarily with excess labor supply (unemployment and underemployment). Finally, equation (5.11) indicates that the initial inequality in the distribution of economic assets among the population determines the *level* of inequality in the personal distribution of national income; around this level, there will be small short-run variations due to changes in the two exogenous variables, money supply and international terms of trade.

The omega model shows, by construction, that the macrobehavior of output and distribution has microfoundations and that, at the same time, the underlying microbehavior of social actors has macrofoundations. The unity of knowledge in the omega model is thus ensured.

Empirical Consistency: Third World Countries

Do Third World countries resemble the omega economy? The answer depends upon whether the beta propositions of this omega model are consistent with the empirical regularities listed in chapter 2.

The omega static model can be confronted against the short-run regularities, which include Facts 2 and 4. The static model indeed predicts Fact 4, the interplay between nominal and real variables in the short run. Regarding Fact 2, the omega model predicts that any general equilibrium situation will imply that the average wage rate is higher than the average income of the self-employed for similar level of skills ($w>v$). This equilibrium condition is observable and makes the model refutable. The model also predicts the existence of unemployment and underemployment. In sum, the empirical predictions of the omega model are consistent with Facts 2 and 4.

Further empirical evidence can be presented now. Consistent with the regularity established under Fact 2, the proportion of workers engaged in wage employment should be higher in the First World compared to the Third World. A sample of 8 First World countries and 10 Latin American countries in the ILO data set showed, indeed, that the ratio of wage employment to total labor force had a mean value of 84 percent in the First World and 59 percent in Latin America around 1996.[1]

According to the omega model, the size of the excess labor supply is equal to the sum of unemployment plus underemployment. The latter is defined

as the self-employed workers earning incomes below the market wage rate, for similar skills. If this gap did not hold in reality, the omega model would fail.

Empirical studies measuring this income gap are not frequent in the international literature. Only two country studies can be cited so far. A study on Peru, based on the national household survey of 2003, found, first, that indeed the wage rate was higher than the mean income of the self-employed among comparable workers in education level: on average, the premium was 55 percent; and, second, that the rate of underemployment was 51 percent and that of unemployment was 7 percent, estimating a rate of total excess labor supply equal to 58 percent (Figueroa 2010, Tables 4 and 5, pp. 124–125). Another study on Peru, also based on national household surveys, estimated that the average premium was 45 percent for the period 2007–2010 (Rodriguez 2013).

The other study is on Brazil. The study is based on a sample of the national census of 1980, in which the sample size was 3 percent and the universe the large cities of Brazil and workers with 11 years of education or less (excluding postsecondary levels). This study finds that the income gap between wage earners and the self-employed is, on average, around 30 percent (Telles 1993, Table 1, p. 239). No estimation of the rate of excess labor supply is reported. Thus, according to the available data, Fact 2 seems to be consistent with the predictions of the omega model.

Conclusions

The economic structure of the omega model consists of the capitalist sector and the subsistence sector. The static general equilibrium of the omega society is determined sequentially: first, the equilibrium in the capitalist subsystem is attained, including wage employment; then output and employment in the subsistence sector is determined; unemployment is a residual of the excess labor supply. The static general equilibrium implies positive rates of unemployment and underemployment, together with inequality, including inequality among homogenous workers.

According to the omega model, labor discipline in the capitalist sector is ensured via underemployment, not via unemployment. Omega society may operate with or without unemployment; it is not a necessity, as it was in the epsilon society. Therefore, underemployment is the necessity for the functioning of capitalism in an overpopulated society. According to the omega model, the peasantry, for example, plays a significant role in making the functioning of capitalism viable in the Third World, for they take care of their own livelihood as farmers, exploiting their marginal lands. If for some

reason their marginal lands became unproductive (curve mn disappears in figure 5.1 panel (c)), then this landless peasantry would increase the size of the unemployed. Capitalism cannot be socially viable in this overpopulated society with such a large redundant labor.

According to the omega model, unemployment and underemployment constitute the excess labor supply in the Third World. The common practice of using the unemployment rate as the only criterion for making international comparisons in the excess labor supply is thus not warranted. It is clear that the excess labor supply in the First World is given by unemployment alone. However, in the Third World unemployment figures largely underestimate the magnitude of the total excess labor supply.

The omega model is consistent with the empirical regularities of Third World countries in the short run. Indeed, the omega model predicts Facts 2 and 4, stated in chapter 2. However, there is another empirical regularity that is also notable in the Third World. This is Fact 3: the existence and persistence of income gaps between ethnic groups. Omega theory makes an abstraction of social differences in society and cannot explain this phenomenon. Omega society thus resembles Third World countries with weak or no colonial legacy. Another theory of the Third World, in which colonial legacy is considered important, will be developed in the following chapter. This theory should predict Fact 3, in addition to Facts 2 and 4, which are general for the Third World countries.

CHAPTER 6

The Sigma Society

In epsilon and omega societies, individuals are homogeneous in every respect, except in their endowments of economic assets. Political entitlements are equally distributed between individuals; hence, these societies are socially homogeneous. Now consider a capitalist society in which political entitlements are unequally distributed leading to first-class and second-class citizens. This chapter will present an abstract heterogeneous capitalist society, called sigma society. This theory aims at explaining production and distribution in the Third World countries that have strong colonial legacy.

Most Third World countries have a European colonial history. They are multiethnic and multicultural today as result of this historical legacy. As shown in chapter 2, the existence and persistence of significant income gaps between ethnic groups is an empirical regularity that has been established by several empirical studies in countries of Latin America, Africa, and Asia. These empirical studies also show that the gaps are systematic, that is, ethnic groups maintain their positions in the income pyramid of their respective countries; in particular, those ethnic groups that are descendants of aboriginal populations or slave populations of the colonial period tend to be the poorest in the income distribution of each country.

The other empirical regularity established in chapter 2 is that in the Third World self-employment is significant and that the average income of the self-employed is lesser than the average wage prevailing in the labor market, for similar level of skills, measured by schooling years. It follows that the situation of the self-employed is involuntary, because they would prefer to have wage employment. Therefore, self-employment is part of the excess labor supply in the Third World, which is defined as underemployment.

The other part is unemployment. These two empirical regularities need a theoretical explanation. Omega theory assumed overpopulated societies, but they were socially homogenous; so it was able to explain the second fact, but not the first. A new theory is then needed. This chapter seeks to provide such theory.

Sigma society is both overpopulated and socially heterogeneous. How would such capitalist society function? Does sigma society resemble well the real world, namely the Third World with strong colonial legacy? A sigma model will then be developed from which refutable empirical predictions can be logically derived. It will be shown that the model indeed predicts the two empirical regularities.

Sigma: Socially Heterogeneous Class Society and Overpopulated

Sigma is an abstract class society. Individuals are endowed with unequal economic assets. However, sigma is also a socially heterogeneous society, for workers are entitled with unequal political assets. Sigma is thus a hierarchical society. It is also an overpopulated society in the sense that the marginal productivity of total labor is near zero.

Conceptually, assets are goods that generate a flow of incomes. They include not only economic assets (land, physical capital, and human capital), but also social assets; therefore, the assumption here is that social assets also generate a flow of incomes. Social factors are thus introduced into the economic process. However, social assets are special goods for they belong to the realm of rights and entitlements granted to individuals in society. They are no physical goods, nor are they marketable.

Social assets basically refer to political and cultural assets. Political assets indicate the capacity of individuals to exercise individual and collective rights, including the right to have rights. Inequality in the individual endowment of political assets generates a hierarchy of citizens in society, first-class and second-class citizens. As a result, not all individuals are equal before the law; more generally, not all individuals have the same degree of access to public goods, in quantity and quality. Public goods are local rather than universal.

Cultural assets in turn refer to the right of social groups to cultural diversity in a multicultural and multiethnic society. Inequality in the endowment of cultural rights generates ethnic groups with a hierarchy of ethnic markers in society: there are first-class and second-class races, languages, religions, and customs. These social markers are called cultural because their hierarchy is socially constructed and transmitted from generation to generation

as well. Inequality in cultural assets leads to social practices of segregation, exclusion, and discrimination against some ethnic groups.

Individual endowments in political and cultural assets are assumed to be highly correlated in sigma society; hence, only unequal political assets—differences in citizenship—will be included in the construction of sigma society. There will exist social classes and underclasses.

The primary assumptions of the sigma theory (σ) are summarized as a set of alpha propositions as follows:

$\alpha(\sigma)$. *(1) Institutional context*: (a) Rules: Individuals participating in the economic process are endowed with economic and political assets. Economic assets are subject to private property rights. Individuals exchange goods subject to the norms of market exchange, which include the norm that nominal wage rates in labor markets cannot fall; the market system operates with Walrasian and non-Walrasian markets, in which the labor market is of the latter type. The political system is democratic but includes formal or informal norms for excluding some social groups from full citizenship; (b) Organizations: households, firms, and the government.

$\alpha(\sigma)$. *(2) Initial conditions*: (a) Initial inequality: individuals are endowed with unequal quantities of economic assets and unequal political entitlements. The social structure includes social classes (capitalists and workers) and classes of citizens (first class and second class). (b) Initial factor endowments: the stock of capital per worker is such that the marginal productivity of the total labor force is near zero, that is, sigma is an overpopulated society.

$\alpha(\sigma)$. *(3) Economic rationality*: Consistent with the institutional context of capitalism, individuals act motivated by self-interest. Capitalists seek two particular objectives, hierarchically ordered: first, maintenance of class position and, second, maximization of profits. In the labor market, workers seek to maximize wages and minimize effort, while capitalists seek to minimize wages and maximize effort. Due to this conflict in labor relations, capitalists use underemployment as the device to extract effort from workers. Individuals have a limited tolerance for inequality.

A Static Model of the Sigma Theory: The Short Run

In order to make sigma theory falsifiable, a sigma model must be constructed, which can be seen as the task of constructing a particular stage set in which social actors play their roles. The following auxiliary assumptions

(which do not contradict the primary assumptions) are introduced for that purpose:

- There are two social classes: capitalists and workers. Among workers, there are two groups: the X-workers and the Z-workers. Compared to X-workers, Z-workers are endowed with lower levels of human capital and lower political entitlements. This difference comes from the legacy of sigma's colonial history, as Z-workers are the descendants of the populations that were the subaltern under the colonial system. Thus, there are three social groups: Z and X among workers, and A, which corresponds to the capitalist class. Groups X and A belong to the same ethnic group. It is assumed that people cannot choose their ethnic identity, that is, ethnicity is exogenously determined.
- On factor endowments, the sigma model assumes that there exists overpopulation of both X-workers and Z-workers. There exists a certain threshold of human capital needed to learn and operate the technology used in the capitalist sector. Z-workers are by assumption endowed with human capital that is below this threshold. However, Z-workers will be utilized in the capitalist sector as nonskilled labor. Therefore, there will be two labor markets, one for the skilled and the other for the unskilled. They operate as non-Walrasian, in which equilibrium takes place with excess labor supply, for each market operates with efficiency wages. The choices of the redundant labor include unemployment and self-employment in small production units in the subsistence sectors.
- Only one good (B) is produced in sigma society. The level of labor productivity is highly differentiated between the three sectors: the capitalist sector shows the highest level, the Z-subsistence sector the lowest, and in between lies that of the X-subsistence sector. This order reflects differences in capital endowments (physical and human) between firms in the capitalist sector and production units in the subsistence sectors; differences of labor productivity between subsistence sectors reflect differences in human labor endowments. Labor productivity levels in the subsistence sectors are too low to pay wages and generate profits.
- Production in the subsistence sectors is for direct consumption by the producer rather than for the market. Economic units in the subsistence sectors are, therefore, production and consumption units at the same time. This assumption is a logical implication of modeling a one-good society. In a several goods model, this assumption would imply that exchange takes place within each subsistence sector, based mostly on the rules of reciprocity exchange, and also between sectors based on the

rules of market exchange. To be sure, this is not to deny the existence of market exchange between the capitalist sector and the noncapitalist sectors in the real world; the assumption of the model is rather that market exchange between sectors is not the essential factor to understand production and distribution in the Third World, and thus it may be ignored; what is essential is labor exchange.

- Five markets constitute the market system: commodity, money, foreign exchange, and two labor markets (skilled and unskilled). Money supplied by the government is used as the means of payment. Capitalist firms seek to exchange goods in international markets, using foreign exchange as the means of payment. Wages are paid at the end of the production period; consequently, workers need to hold cash balances for transaction purposes. All markets operate under perfect competition.
- The economic process under analysis is static and open.

Behavior of Capitalist Firms

Sigma society exports good B and imports good C, which is a material input required to produce good B. Hence, the domestic price of commodities B and C is given by

$$P_b = P_e P_b^* \qquad (6.1)$$
$$P_c = P_e P_c^*.$$

P_e is the nominal exchange rate (the domestic nominal price of foreign exchange) and P_b^* and P_c^* are international prices denominated in units of the foreign exchange. Under perfect competition, each domestic price must be equal to the international price multiplied by the nominal exchange rate. If this equality did not hold, people could buy commodities from the cheaper source and sell it wherever it is more expensive and make a profit. The ratio $z^* = P_b^*/P_c^*$ is the *international terms of trade*.

It should be clear that the domestic price level P_b refers to the good produced domestically (good B), which depends upon the exchange rate and the international price of good B. Given the value of the latter, the domestic price level varies according to the variation of the exchange rate. Price stability implies exchange rate stability.

Firms seek profit maximization. Nominal profit is equal to the difference between total revenue minus total cost; total revenue is equal to price times quantity of good B sold in the market; total cost is equal to the sum of the wage bill (nominal wage times quantity of workers hired) and the total cost of material inputs (domestic price times quantities of good C imported).

Production technology is such that output of good B depends on the services of the stock of physical capital endowed by each firm, the quantity of workers hired, and the material input C bought abroad. Assume that input C enters into the production of good B in fixed quantities per unit of output, that is, assume that input C is not substitutable for capital or labor, but capital and labor are substitutable to each other. Then for every unit of output, the firm must pay a given amount for input C, which is required for technological reasons for the production of good B. Therefore, the *net* value of average labor productivity is now the relevant concept, which is equal to total net value of output (net of the cost of input C) divided by the number of workers; similarly the *net* value of marginal productivity of labor is the relevant concept, which is equal to the additional *net* value of output that results from hiring an additional worker.

The behavior of the individual firm j can then be represented by the following structural equations:

$$\text{Maximize } P_j' = P_b\, Q_{bj} - P_h\, D_{hj} - P_c\, D_{cj} \tag{6.2}$$

subject to the production function constraints

$$Q_{bj} = \lambda\Phi_j\,(D_{hj}, K_{bj}), \text{ such that } \lambda = \lambda^* = 1$$
$$D_{cj} = c\, Q_{bj}$$
$$X_{bj} = (1/z^*)\, D_{cj}$$

and subject to efficiency wages, that is,

$$\lambda = \lambda^* = 1 \text{ requires that } w > w^w.$$

Profits (P′) refer to nominal values. Total output is net of depreciation, so is total profits. The term c is a technological coefficient, equal to the number of units of good C required to produce one unit of good B. The two first constraints assume that the production function is subject to *limitational factors* and is thus represented as a system of equations. Because good C is an imported input, the firm pays foreign firms with foreign exchange that obtains by exporting part of its output (X_b) according to the terms of international trade z^*. Labor refers to workers only. The last constraint indicates that efficiency wages must prevail in the labor market, which implies that market wage rate must be higher than the Walrasian wage rate (w^w) in order to maintain the required labor discipline and thus the productivity level.

The structural equations presented above lead to the following equilibrium condition:

$$\Phi_{j1}\,(D_{hj}, K_{bj})\,(P_b - c\,P_c) = P_h, \text{ or} \tag{6.3}$$
$$\Phi_{j1}\,(D_{hj}, K_{bj})\,(1 - c/z^*) = w, \text{ where } (c/z^*) < 1.$$

The marginal productivity of labor in the capitalist sector is now net of the cost of using the limitational input C. Profit maximization implies equality between the value of the net marginal productivity of labor and the nominal wage rate or, alternatively, equality between the net physical marginal productivity of labor and the real wage rate. Total profits arise from the difference between the average productivity of labor and the marginal productivity of labor multiplied by the employment level.

Individual firms are price takers in all markets. Therefore, for each firm, profit maximization implies hiring workers up to the point that the value of the net marginal productivity of labor is equal to the market nominal wage rate. Given the exogenous variables, firms will choose to produce a quantity of good B, buying a quantity of input C and hiring a quantity of workers. As long as the exogenous variables remain fixed, these quantities of the endogenous variables will be repeated period after period. This is the static equilibrium in the behavior of the typical firm.

Now introduce unskilled labor or Z-workers. The assumption is that unskilled labor is required to do supplementary activities in production and in fixed proportion to the quantity of skilled workers. The equilibrium condition is still that the marginal revenue must be equal to the marginal costs in units of labor, that is, the net marginal productivity of skilled labor must be equal to the cost of hiring an additional skilled worker, which now includes the market real wage rate of the skilled worker (w_x) plus the quantity of unskilled labor need for each skilled worker (ε) multiplied by the wage rate of unskilled worker (w_z), that is $P_h = P_{hx} + \varepsilon P_{hz}$ and $w = w_x + \varepsilon w_z$. According to the *composite good theorem*, a group of goods can be treated as a single good if they are utilized in fixed proportions. We can then treat skilled labor as a composite factor with price equal to w (not to w_x), as long as coefficient ε is kept fixed.

The behavior of firms is easily derived. From the equilibrium condition of a typical firm (which is stable), the effect of change in the exogenous variables upon the endogenous variables can be derived through comparative statics. The relevant endogenous variable is the quantities of labor demanded; whereas the exogenous variables include the price level (P_b), the nominal wage rate (P_h), the terms of international trade (z^*), and the capital stock (K_b). Aggregating the individual labor demand of firms over

all firms, the market demand for labor can be represented by the following function:

$$D_b = H(P_b, P_c, P_h, K_b),\ H_1>0,\ H_2<0,\ H_3<0,\ H_4>0 \tag{6.4}$$

subject to budget constraints

$$P_b S_b = (P_c D_c - P_b X_b) + P_h D_h$$
$$X_b = (1/z^*) D_c = (1/z^*) f(D_h, K_b),\ f_i>0.$$

Behavior of Workers

Workers seek to maximize consumption or real income, which implies seeking employment in the capitalist sector. The behavior of workers who are wage earners is then easy to derive in this model. They seek to maximize real income (D_b) subject to their individual budget constraints and to the requirement of holding cash balances for transaction purposes (D_m). Thus for the individual worker j, this assumption can be written as

$$P_h D_{hj} = P_b D_{bj} + (D_{mj} - S_{mj}) \tag{6.5}$$
$$D_{mj} = P_b L(D_{bj}),\ L'>0,\ L''<0.$$

The equilibrium condition is that the money supply should be willingly held, which implies that total wage income is spent in buying good B. Note that workers cannot decide on the quantity of labor services that they want to sell in the labor market because the labor market is non-Walrasian.

The static equilibrium is (trivially) stable. Therefore, the comparative statics method can be applied to derive the effects of exogenous variables upon endogenous variables. Aggregating over all workers, we get the following reduced form equations:

$$D_b = F(P_b, D_h, P_h),\ F_1<0,\ F_2>0,\ F_3>0 \tag{6.6}$$
$$D_m = P_b L(D_b),\ L'>0,\ L''<0$$

subject to the aggregate budget constraint of wage earners

$$P_h D_h = P_b D_b + (D_m - S_m)$$

As a first option, workers seek employment in the capitalist sector, for the market wage rate is higher than the income they can make in the

subsistence sector. The incentive system leads them to this endeavor, but jobs in the capitalist sector are not available for all workers, no matter how hard each worker tries. Because there is a turnover of workers in the capitalist sector due to dismissals of those workers found shirking, there is a probability of finding a job. Assume this probability is the same for all workers.

Let w^e represent the workers' expected wage if looking for a job in the capitalist sector. For an exogenously given probability π of finding a job, the model assumes that the expected wage depends positively on the market wage rate and is thus made observable. These assumptions imply a uniform expected wage for all workers, which can be written as

$$w^e = \pi w, \text{ where } 0 < \pi < 1 \tag{6.7}$$

As a second best solution, the workers who become excluded from employment in the capitalist sector will choose between unemployment and self-employment. These workers will thus evaluate the expected wage (what they would get finding a job after the time spent looking for it) against the sure income they can make in the subsistence sector.

For a given market real wage rate, therefore, the redundant workers will face an expected income if they choose to be unemployed. On the other hand, the higher the number of workers who choose self-employment, the lower the income they make, due to diminishing returns; therefore, for a number of these workers the expected wage rate would be lower than that they can make in the subsistence sector, but for others it would be higher. Assuming that redundant workers are homogeneous and have no preferences in favor of any of the two residual options, the equilibrium allocation of workers between unemployment and self-employment will take place when the expected wage rate is equal to the sure marginal productivity of labor in the subsistence sector.

Hence, a group of the excluded workers would choose to seek jobs in the labor market, whereas the rest would choose self-employment in the subsistence sector. The group who chooses to seek a job becomes *voluntarily* unemployed, as self-employment is an open alternative. However, note that this is a choice under second best alternatives. The first option, the workers' most preferred situation, is to have a job in the capitalist sector. The model assumes transitions of workers between employment in the capitalist sector, unemployment, and self-employment in the subsistence sectors, whenever jobs are increasing or decreasing in the capitalist sector.

General Equilibrium

There are four markets that operate under perfect competition in the capitalist sector. The conditions of equilibrium for each market are derived from equations (4) and (6), as follows:

$$\text{Labor market } S_b \equiv D_b + E_b,\ w>w^w \tag{6.8}$$
$$\text{Commodity market } S_b = D_b$$
$$\text{Foreign exchange market } P_c^*\ D_c = P_b^*\ X_b$$
$$\text{Money market } S_m = D_m$$

subject to the aggregate budget constraint

$$P_b\ (D_b - S_b) + P_e\ (P_c^*\ D_c - P_b^*\ X_b) + (D_m - S_m) = 0.$$

The labor market is non-Walrasian; it operates with excess labor supply (E_b>0) and is subject to the constraints that the nominal wage cannot fall and the market wage rate must be higher than the Walrasian wage rate. The only labor considered is that of skilled labor (X-workers). The quantity demanded of Z-workers is determined once the employment of X-workers is known. The rest are Walrasian markets. The market system will solve for the values of the endogenous variables, for given values of the exogenous variables.

We now need to determine the core of the general equilibrium. Equilibrium in the foreign exchange market can be incorporated into the labor market equation. The reason is that at any given value of D_b there will correspond a quantity of total output, which will require imported inputs and foreign exchange in certain amount, which will be paid for by exporting part of the output, leaving a quantity of net output for the domestic market, and the foreign exchange market in equilibrium. This is true for any equilibrium value of employment in the labor market. The nominal exchange rate will be determined in the money market.

Therefore we are left with two Walrasian markets (good B and money), of which one can be eliminated by invoking Walras's law. Thus, general equilibrium can be established by solving two markets, say the money market and the labor market. Then the core of the general equilibrium system is given by the following equations:

$$D_b = H\ (P_e;\ P_b,\ z^*,\ K_b),\ H_1>0,\ H_2<0,\ H_3>0,\ H_4>0 \tag{6.9}$$
$$S_m = P_e\ \ P_b^*\ \ L\ (W) = M\ (P_e,\ D_b;\ P_b,\ \ P_b^*\ ),\ \text{where } M_i>0,\ \text{for all } i$$
$$\text{subject to } w \geq w^* = [(1+\rho)\ v'] > w^w.$$

The core is constituted by two equations and two endogenous variables, the exchange rate P_e and the employment level D_h. The solution is simultaneous. Once these values are known, the rest of the endogenous variables can be solved just by implications; thus employment level determines total output and exchange rate determines the price level (P_h), which in turn determines the real wage rate ($w=P_h/P_h$). The constraint says that equilibrium real wage rate must be equal to or higher than the marginal productivity of labor plus a given premium rate ρ, and in any case higher than the Walrasian wage rate.

Once the values of equilibrium in the X-labor market have been determined (D_h and w), the allocation of the excess labor supply into unemployment and self-employment will be solved. The relevant equations to solve this allocation include

$$E^0_{hx} = S_{hx} - D_{hx}{}^0 = L_{sx} + U_x \tag{6.10}$$
$$V_x = J(L_{sx}), \text{ where } J'>0, J''<0$$
$$\pi w_x = J'(L_{sx}) \equiv v_x', \ 0<\pi<1 \text{ and } \pi w^0{}_x > J'(E^0_{hx}).$$

The first equation points out that in the sigma economy excess labor supply of X-workers takes the form of unemployment and self-employment. The second equation is the production function in the subsistence sector, which is subject to Ricardian diminishing returns. These two are the structural equations. The third equation shows the equilibrium condition for the allocation of labor to self-employment: workers will seek to equalize the marginal productivity of labor in the X-subsistence sector (v_x') to the expected wage rate (πw_x).

Given the curve of the marginal productivity of labor in the X-subsistence sector, the self-employment level will be determined by the expected wage rate; given the total amount of excess labor supply, the unemployment level of equilibrium will be determined once the self-employment level is known. Thus, the unemployment level is a residual of the residual labor; it could even be zero in equilibrium. The allocation of excess labor supply in the Z-labor market will also be allocated to self-employment and unemployment under similar principles to the ones shown for X-workers.

Figure 6.1 illustrates the general equilibrium situation. Panel (a) depicts the X-labor market. The quantity supplied of X-workers is equal to the segment OO′. The curve HR represents the marginal productivity of labor in the capitalist sector. Curve mn (measured from origin O′) is the marginal productivity of labor in the subsistence sector. Curve m*n* is the expansion of curve mn by the factor ρ (say 50 percent), the premium or gap needed

to maintain the labor productivity level; this is called the *effort extraction curve*. Any wage rate that implies a smaller premium rate (say 20 percent) will bring the level of the productivity curve down and reduce profits, and this will be no solution. At value $\rho=0$, we would have the Walrasian wage rate, which would imply full employment at point N, but this is also unattainable. Thus, only the segment Hg of the marginal productivity of labor is the relevant labor demand curve.

Let the equilibrium real wage rates be w_x and w_z, such that the total cost of employing a unit of X-worker is equal to $w=w_x+\varepsilon w_z$. Given the labor demand curve Hg, the equilibrium in the capitalist sector is given by point E, which implies wage employment OA. Total output is equal to the area under segment HE, which is distributed as profits (P), total wage bill of X-workers (W_x), and total wage bill of Z-workers (W_z). Thus total self-employment is equal to segment BO′, which implies a level of unemployment equal to segment AB.

Panel (b) shows the equilibrium situation in the Z-subsistence sector. Total quantity of Z-workers is given by the segment O′O″. Curves rs and r*s* (measured from origin O″) represent the marginal productivity of labor and the effort extraction curve. At the equilibrium real wage rate w_z, employment is endogenously determined (as the required complementary labor consistent with the quantity of X-wage employment), and is equal to segment O′A′. It should be noted that the area of total wage bill of Z-workers (W_z) must be equal to the corresponding area shown in panel (a). Then it follows that A′O″ is the excess labor supply of Z-workers. The economic choice of Z-workers between self-employment and unemployment is guided by the same rationality of X-workers, shown in panel (a), that is, the expected wage rate (πw_z) must be equal to the marginal productivity of labor (v_z'); therefore, equilibrium self-employment will be equal to segment B′O″; hence, A′B′ is equilibrium unemployment.

The two labor markets are clearly non-Walrasian. Real wage rates do not adjust excess demand or excess supply, except that they must correspond to the set of efficiency wages. This is the case in both labor markets. In the first labor market, the real wage rate w_x lies above the effort extraction curve, which will maintain the productivity level (segment Hg) fixed. Point g indicates the *effective full employment* of X-workers, that is, the maximum level of employment together with the minimum real wage rate of efficiency (w^*). If for some reason, the market real wage rate was the Walrasian price, implying full employment, point N, the curve HR would shift downward; hence, point N is unattainable. In the second labor market, the real wage rate w_z also lies above the effort extraction curve, which will also maintain the segment Hg fixed in the capitalist sector. Hence, w_x and w_z are equilibrium real wage rates.

Figure 6.1 shows that the static general equilibrium solution is sequential. First, production and distribution of equilibrium in the capitalist sector is determined, which in turn determines the excess labor supply as well; then, equilibrium values of self-employment and total output in each subsistence sector is determined. Unemployment is residual of residual. In this sigma model unemployment is voluntary in the sense that it is the result of the choice of workers, but it is involuntary in the sense that the most preferred situation is to have wage employment and that unemployment is a second best choice: The unemployed have decided to choose searching for jobs in the capitalist sector instead of taking an occupation readily available as self-employed, which however would yield a lower income than the expected wage. Thus, Robert Lucas's (1978) famous proposition that any unemployment is voluntary because "the unemployed worker at any time can always find some job at once" needs a subsistence sector; hence, it applies to sigma (and omega) societies, not to epsilon, as he mistakenly attributed (see chapter 3).

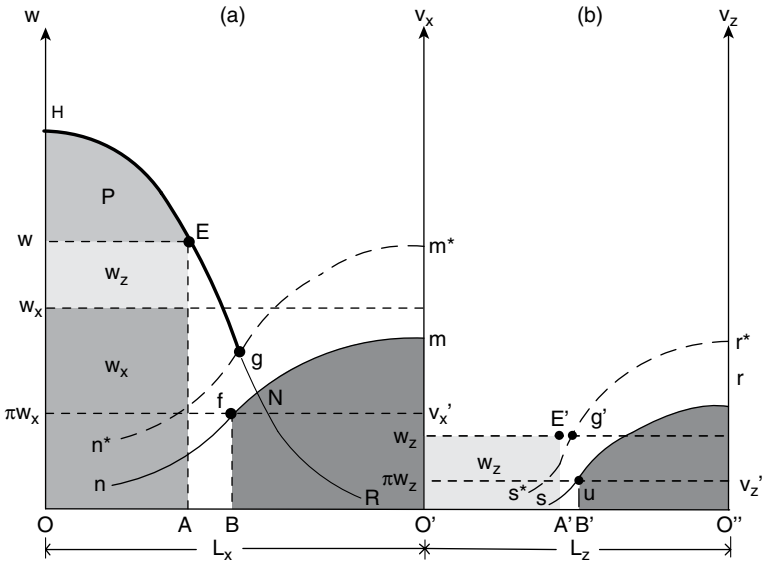


Figure 6.1 General equilibrium in sigma society.

Note: $w = w_x + \varepsilon w_z$.

It can also be seen in the graph that the good B market is also in equilibrium: the quantity supplied to the domestic market is equal to the area under curve HE minus the area of profits, which is just equal to the quantity demanded, which comes from workers alone, and is given by the area including the wage bills of X- and Z-workers. The foreign exchange market is also in equilibrium, for at the equilibrium, value of D_h will correspond to the quantity of total output; this requires imported inputs and foreign exchange in certain amount, which is paid for by exporting part of the output, leaving behind domestically the net output given by the area under HE.

Figure 6.1 also shows that national income is the sum of three areas: the area under segment HE (capitalist sector), the area under segment mf (X-subsistence sector), and the area under segment ru (Z-subsistence sector). Because only one good is produced (good B), this addition is fully justified. National income (Y) of equilibrium and its distribution between social groups (where X-workers and Z-workers are explicitly distinguished by the subindices) can then be represented by the following system of equations:

$$\begin{aligned} Y &= P + W_x + W_z + V_x + V_z \\ &= p\,L_a + w_x\,D_{hx} + w_z\,D_{hz} + v_x\,L_x + v_z\,L_z \\ S_{hx} &= D_{hx} + L_x + U_x \\ S_{hz} &= D_{hz} + L_z + U_z. \end{aligned} \tag{6.11}$$

The first equation of this system indicates that national income is equal to the total output produced in the capitalist sector and the two subsistence sectors. The second equation just shows that national income can be decomposed into average incomes multiplied by the total employment in each sector; total profits P is also decomposed into average profit per capitalist (p) and the size of the capitalist class (L_a). The last two equations show the allocation of labor supply to the sectors and the residual unemployment.

The relation among mean incomes of the different social groups under static general equilibrium can also be derived from figure 6.1. They are

$$\begin{aligned} &\text{Capitalist sector inequality: } p > w_x > w_z \\ &\text{X-labor inequality: } w_x > v_x \\ &\text{Z-labor inequality: } w_z > v_z \\ &\text{Between subsistence sectors inequality: } v_x > v_z \\ &\text{Average income by ethnic groups: } p > y_x > y_z. \end{aligned} \tag{6.12}$$

The first inequality shows that the mean income of capitalists is higher than the real wage rate of X-workers, which in turn is higher than that of Z-workers. This is the result of the initial inequality in the distribution of physical and human capital. The second and the third indicate the result of the labor discipline device: market real wage rate must be higher than the opportunity cost of wage earners, the average productivity of labor in the subsistence sector. The fourth inequality also corresponds to differences in factor endowments in human capital: the level of output per worker is higher in the X-subsistence sector compared to that in the Z-subsistence. The last inequality shows the aggregate average income by ethnic groups: Z-workers constitute the poorest social group in sigma society.

According to this sigma model, the two subsistence sectors play a crucial role to make viable the functioning of capitalism in an overpopulated society. Redundant workers are responsible for their livelihood as self-employed workers. If the subsistence sectors did not exist, all self-employed workers would become unemployed. Part of profits and wages would have to be transferred to the excess supply of workers; hence, with such a large unemployment rate, the capitalist sector would not be socially viable. Society would function under another system.

X-workers could set up firms because they are endowed with the similar human capital as those workers employed in the capitalist sector. According to the model, they cannot. They cannot rent capital to open a firm, for there is no market for capital service, as capitalists prefer profits to rents to retain their class position. Then workers would have to invest in physical capital, which in turn needs financing. Nevertheless, these workers (aiming at having small businesses) will be excluded from both the credit market and the insurance market making these projects unviable. Banks and insurance firms find it unprofitable to exchange financial services with small businesses due to the high transaction costs.

Residual Z-workers will face even stronger limitations to escape from their situation of poverty. Renting or buying physical capital will be unviable, as they are constrained by their low human capital endowments to use capitalist technology and machinery. Exclusion from credit and insurance markets will be even stronger.

In sum, the core of the general equilibrium has a solution, a unique one. Moreover, this is indeed an equilibrium situation: no one has both the power and the will to change it. In this static general equilibrium situation, the values of the endogenous variables will be repeated period after period as long as the values of the exogenous variables remain fixed.

Beta Propositions

The general equilibrium of the sigma model is stable. The labor markets are non-Walrasian and the money market is Walrasian and stable. Then empirical propositions can be derived from the model by the comparative statics method. In the short run, the relevant exogenous variables are money supply and international terms of trade. The effect of changes in these exogenous variables upon national income and its distribution can easily be resolved in this very simple model.

Consider first an increase in the money supply. This will generate directly an excess of cash balances willingly held by workers, who will then intend to restore the equilibrium by using this excess to buy more quantities of good B in the period of adjustment. This adjustment would reduce both the exports and foreign exchange inflows and thus an increase in the nominal exchange rate. Then the price level would increase and the real wage would fall; hence, firms would seek to hire more workers (from the pool of excess supply) and produce more output to obtain more profits. This would be the first round effect. The second round effect will be ambiguous because the fall in the real wage rate and the increase of employment imply an ambiguous change in the real wage bill, which in turn implies an ambiguous and small change in the demand for real cash balances and on the price level. Therefore, the second round effect upon the price level would be small and may be neglected. The effect of an increase in money supply is to decrease the two real wage rates and to increase employment of X-workers and Z-workers in the capitalist sector.

Figure 6.1 panel (a) illustrates the effect of an expansionary monetary policy upon equilibrium in the capitalist sector, where point E represents the initial general equilibrium. An increase in the money supply increases the price level, which implies a fall in the real wage rate and an increase in the employment level. The initial equilibrium situation, at point E, will be moved to another equilibrium situation, which will lie below point E, along the segment Eg. The government can then reduce the excess supply of labor using monetary policy.

A sufficiently large increase in money supply will ultimately reach the effective full employment level, at point g. If money supply keeps increasing, employment will remain fixed, and only the exchange rate and consequently the price level will rise, which will then induce increases in the nominal wage rate in the same proportion in order to avoid the fall in real wage below the minimum efficiency real wage rate (w^*), that is, *money becomes neutral*. It should be noticed that full employment equilibrium—the Walrasian market equilibrium—given by point N is unattainable in the sigma model.

On the subsistence sectors, the consequences are as follows. The increase in employment in the capitalist sector implies a fall in the excess labor supply of both types of labor. The fall in the real wage rates implies a fall in the expected wage rate, which in turn implies that more workers who are redundant will choose self-employment. In each subsistence sector, higher self-employment will increase total output, but marginal productivity of labor will fall. Unemployment of each type of workers will shrink on two accounts: excess labor supply falls and self-employment increases.

On income distribution, the effect of an increase in money supply is as follows. With respect to the initial general equilibrium, at point E, at point f, and point u, the new production equilibrium shown earlier implies higher total output both in the capitalist sector and in the two subsistence sectors. Changes in their shares in national income are ambiguous. Within the capitalist sector, total profits increase, but the change in the wage bill is ambiguous; hence, change in the share of wages and that of profits in total output is ambiguous. Change in functional income distribution is thus undetermined.

As to changes in personal income distribution, average profit of capitalists increases in absolute and relative terms, while real wage rates fall, and average real incomes in the subsistence sector fall as well. Relative personal income of the capitalist class increases relative to that of workers, but changes in relative incomes among workers is ambiguous; at the same time, the proportion of the unemployed falls. Hence, Lorenz curves will tend to cross each other and overall income inequality will tend to remain unchanged, or changes will be small, that is, the Gini index will remain unchanged or the change will be small. What is clear though is that the order of average incomes among ethnic groups—system (6.12)—will remain unchanged.

Now consider the effects of an increase in the international terms of trade. As can be visualized in figure 6.1 panel (a), an increase in the international terms of trade (z^*) shifts the marginal labor productivity curve HR outward. The initial general equilibrium is at point E. Firms would seek to hire more workers drawn from the pool of excess supply; hence, the market real wage rate could remain unchanged. However, this is not the case. The wage bill will initially increase. Given the quantity of money supplied, the buildup of cash balances would imply a temporary fall in the quantity of good B demanded, which would increase both exports and the inflow of foreign exchange; this would lead to a fall in the exchange rate, which in turn would lead to a fall in the domestic price level and a rise in the real wage rate; hence employment would still increase, but not as much as indicated in the initial effect. Then the initial equilibrium situation will be moved to another equilibrium situation, say point E″ (not marked), which will be located to

the northeast of point E, where both employment and real wage rate in the two labor markets will be higher.

The increase in the international terms of trade implies an increase in both the real wage rates and employment, which in turn implies a reduction in the excess labor supply of both types of labor. However, the increase in the real wage rates leads to higher expected wage rates and thus to a decrease in self-employment, and thus to a fall in total output in the two subsistence sectors. Change in unemployment is ambiguous because the effect of higher wage employment, maintaining the real wage rate constant, is to reduce unemployment; but the increase in real wage rate produces the effect of increasing it, as self-employment falls in each subsistence sector.

Figure 6.1 allows us to visualize the effect of an increase in the international terms of trade upon income inequality. Total output increases in the capitalist sector, but it declines in each subsistence sector. Within the capitalist sector, the wage bill increases, but the change in profits is ambiguous due to the increase in the real wage rate; hence, the share of profits in national income is ambiguous, and so is the share of wages. Changes in functional income distribution are undetermined.

As to changes in relative incomes, the change in the average income from profits in the capitalist class is ambiguous, for real wage rate goes up. The real wage rate and the average income from self-employment both increase, but changes in the relative incomes are ambiguous. What is clear is that the order of average incomes among ethnic groups—system (6.12)—will remain unchanged. This result, together with that of money supply, indicates that the inequalities system (6.12) constitutes another equilibrium condition of the short-run sigma model.

The reduced form equations—which show the causality relations of the model—for total output or national income (Y), quantity of labor demanded in the capitalist sector (D_h), and income inequality (D) in sigma society can then be written as follows:

$$Y^0 = F^\sigma (S_m, z^*; \delta, K_h, S_h) \tag{6.13}$$
$$F_3 = 0 \text{ and } F_i > 0 \text{ otherwise}$$
$$D_h{}^0 = G^\sigma (S_m, z^*; \delta, K_h, S_h) \leq D_h{}^* < S_h$$
$$G_3 = 0 \text{ and } G_i > 0 \text{ otherwise}$$
$$D^0 = H^\sigma (S_m, z^*; \delta, K_h, S_h)$$
$$H_3 > 0 \text{ and } H_i = (?) \text{ otherwise.}$$

These results come from solving the core of the general equilibrium, which is shown in the Appendix, and then adding their implications for these three endogenous variables.

The first equation indicates that factor endowments (and technology) determine the *level* of national income in the long run. A higher factor endowment will imply a higher output level. Given the factor endowments, there will be short-run variations around the national income level due to the effect of changes in the two exogenous variables, money supply and international terms of trade. The second equation shows the equilibrium quantity of employment in the capitalist sector, which cannot take values equal to the quantity supplied of workers (X-workers and Z-workers combined); hence, the labor market equilibrium is with excess labor supply, as the quantity of labor demanded has a limit (the effective full employment, D_h^*), which is smaller than total labor supply (S_h).

The last equation indicates that the initial inequality in the distribution of economic and political assets among the population (δ) determines the *level* of inequality in the distribution of national income in the long run. The higher the inequality in the distribution of the *stock* of economic assets and political entitlements, the higher will be the inequality in the distribution of the *flow* of incomes. Given the initial inequality, there will be short-run variations around the income inequality level due to changes in the two exogenous variables, money supply and international terms of trade; but the direction of these variations is undetermined and they tend to be small. The implication is that the order of average incomes among ethnic groups will tend to persist in the short run and even in the long run, as long as the initial inequality remains unchanged.

Some additional comments are in order. First, it should be noted that this sigma model predicts that output expansion in the capitalist sector has a "trickle down" effect on the two subsistence sectors, through the labor market. The subsistence sectors are residual to the workings of the capitalist sector. This is not to deny that some relations through commodity markets may exist among the capitalist and subsistence sectors in the real world, but the model assumes that this effect is not significant. The sigma model assumes that the essential mechanism in the relations between sectors operates through the labor market. It is the workings of the labor market that generate the existence of the economic structure—capitalist sector and subsistence sectors—of sigma societies. Unemployment is another outcome of the economic process, but it is not a necessity for the functioning of the sigma society; self-employment plays that role.

Second, general equilibrium assumes that the curve showing the marginal productivity of labor in the X-subsistence sector is *given*. If this curve shifted upward exogenously, then the effort extraction curve would be shifted upward, and the value of w^* would increase. The solution of price and quantities in the capitalist sector would be different. This is to indicate that the labor productivity in the subsistence sector plays a role in the

general equilibrium solution. The sigma model does not predict economic dualism between capitalist and subsistence sectors.

Third, the real wage rate increase in an overpopulated society would seem a paradox. The sigma model has explained the increases in the short run. How about in the long run? To recall, the Lewis model predicted output expansion in the capitalist sector at fixed real wage rates, that is, supply of labor was a horizontal curve. The real wage rate increase required the elimination of overpopulation. However, the sigma model assumes an increasing labor supply curve (curve mn in figure 6.1 panel (a)), which then predicts that capitalist expansion (shifts in labor demand) will cause real wage rate increases before overpopulation has been eliminated.

The other difference with the Lewis model is that the sigma model assumes that in labor markets the real wage rate and the employment level are not determined by real variables alone (factor endowments, physical labor productivity, technology, real opportunity cost) but by nominal variables as well. The real wage rate of equilibrium is determined once the price level is determined, for the nominal wage rate is exogenously given, that is, the role of the real wage rate is not to clear the labor market. Labor markets are non-Walrasian.

Empirical Consistency: Third World Countries with Strong Colonial Legacy

Sigma theory seeks to explain the production and distribution process in those capitalist countries that have a significant European colonial legacy. European colonial systems established different types of institutions in their colonies (Acemoglu, Johnson, & Robinson 2001). An important common institution, rarely mentioned, was political inequality: the (informal) norm of the coexistence of first-class citizens (the colonizer) and second-class citizens (the colonized and the slaved brought from abroad). Historians implicitly recognize this legacy by pointing out the nature of the colonial institutions: "Colonial societies were generally characterized by apartheid and segregation and often were based on notions of innate racial inequalities" (Wesseling 2004, pp. 242–243). Sigma theory assumes that the norm about the inequality in citizenship has prevailed under capitalism.

The sigma model predicts that the equilibrium conditions of production and distribution in the short run will be with income gaps among workers of similar human capital endowment, depending upon whether they are wage earners or self-employed: market wage rate will be higher than average income from self-employment. Therefore, the self-employed are part of the excess labor supply. This prediction is consistent with Fact 2 listed in chapter 2.

In sigma societies, as in the case of omega societies, excess labor supply is a necessity for the functioning of capitalism: it is the device for labor effort extraction, which leads to high levels of labor productivity, and thus to high levels of profits. Full employment equilibrium is unviable, for it cannot lead to those results. The mechanism of labor effort extraction in a sigma society is not unemployment, but underemployment: market wage rate must be higher than the opportunity cost of wage earners, which is given by the marginal productivity in the subsistence sector. Thus, this device mechanism generates both inequality among homogeneous workers and a large subsistence sector. The excess labor supply takes the form of unemployment in the First World, whereas in the Third World it is composed of unemployment and underemployment, the latter being the more significant. The common practice of using the unemployment rate as the criterion for making international comparisons about the excess labor supply is thus unwarranted. Unemployment figures underestimate the magnitude of the total excess labor supply in Third World countries.

The sigma model predicts that in equilibrium relative incomes will vary among workers, depending upon their ethnic origin, and that the Z-populations constitute the poorest groups of sigma society. This prediction is consistent with Fact 3 listed in chapter 2.

Sigma society is a multiethnic and a hierarchical society, a legacy of its colonial history. The endowment of human capital among workers in this society is not a random mechanism. Given the power structure in society, in which political power entitlements are unequally distributed, the human accumulation process is also unequal, as the education system is not human capital equalizing. Z-workers, who are second- rate citizens, get second-rate education, in quantity and quality. So Z-workers always have low-level human capital endowment and vice-versa: those with low human capital endowments are mostly the Z-population.

Facts about differences in human capital endowments among ethnic groups are scarce. In the case of Peru, a study based on the national household survey of 2003 found that indeed education levels (as indicator of human capital) varied by ethnic groups, and that these differences were statistically significant. The mean years of schooling among adult population were 7.6 for Z-population (indigenous population), 11.4 for X-population (mestizo population), and 14.2 for the white population; mean income of the X-workers were two times that of Z-workers (Figueroa 2010, Table 3, p. 122). It is remarkable that after almost 200 years of independence, the gap of human capital levels is significant. This gap, it should be said, is more than what the schooling years indicate, for the quality of education of Z-workers is lower, as will be shown in chapter 2, volume 2.

According to the sigma model, the level of income inequality in Third World countries is determined by the initial inequality in the distribution of economic and political assets. The initial inequality is related to the colonial legacy of these countries. The descendants of the dominated population (native and slaves brought from Africa), called here the Z-populations, are poorly endowed with economic resources and political entitlements compared to the European descendants and the mestizos. In sum, the underlying factor that explains the observed systematic poverty of Z-workers is the colonial legacy. There exists path dependence in sigma societies.

Also, according to the sigma model, the role of the subsistence sectors in the Third World is to make the functioning of the capitalist system viable in overpopulated societies. If the subsistence sectors did not exist, the excess labor supply would become unemployed. To make capitalism viable, a part of profits would have to be transferred to the unemployed. However, this transfer could exhaust profits. Capitalism cannot operate in such a society: the rule of production and distribution cannot be based on the *marginal* productivity of labor (which is equal to zero for total workers), which is the rule for profit maximization (provided the excess labor supply finds ways to satisfy their own livelihood). An overpopulated society could operate under a different rule, say, the one based on the *average* labor productivity, as is the case of feudal societies (Georgescu-Roegen 1960). Therefore, the existence of the subsistence sectors in the Third World is a necessity for the viability of capitalism. Indigenous peasants farming in marginal lands in the Andes, for example, contribute to the viability of capitalism in the Andean countries.

Other predictions of the short-run sigma model include the proposition that the effect of the terms of international trade are positive on national income but ambiguous on income inequality. The effect of an expansionary monetary policy is positive on national income but ambiguous on income inequality. Therefore, income inequality in the short run tends to show some viscosity. No empirical studies are available on these relations.

In the international literature, the observed self-employment in the Third World is usually seen as the informal sector, illegal sector, or the shadow economy. According to this view, this sector is the outcome of economic choice made by individuals, who take into account the cost and benefits of illegality. It is more profitable to act illegally. The excessive cost of legality over its benefits leads people to act outside the legal system, which would be the case with small firms and the self-employed (cf. Schneider & Enste 2000). According to sigma theory, in contrast, the observed self-employment

in the Third World is explained by the workings of capitalism in an overpopulated society. Self-employment is mostly equal to underemployment, which is a form of excess labor supply generated by the functioning of the labor markets in overpopulated and socially heterogeneous societies. This is the essential factor. Just to make sure, the cost of legality does exist in the real world and, surely, it has an effect, though a small one, upon the size of self-employment. This is what the sigma theory says, as it made an abstraction of this factor and its predictions are consistent with facts. Thus, according to the sigma model, if the costs of illegality were eliminated, the size of the subsistence sectors would be reduced, but not by much; the subsistence sectors would still remain. Workers unable to find wage employment would seek self-employment in legal or illegal activities, which is a second-order choice.

Another usual argument to explain informality is that governments are too bureaucratic and corrupt. However, why are governments so? In an overpopulated society, governments have incentives to buy votes by creating public employment. Hence, according to the sigma model, bureaucracy size is endogenous. The size of corruption is also endogenous: in a very unequal society, the state is endogenously weak, and law enforcement is also weak, as will be shown in chapter 7.

Conclusions

In order to explain the functioning of Third World countries with colonial legacy, the large majority of countries, this chapter has presented a new economic theory called sigma theory. The abstract sigma society is overpopulated and socially hierarchical, a legacy of colonial history. To make this theory refutable, a sigma model has been developed from which a set of refutable empirical predictions have been logically derived. The model predicts the two empirical regularities that arise from international studies: the existence and persistence of income gaps between ethnic groups and the significance of self-employment as excess labor supply. The sigma society has a marked resemblance to the real world. The underlying factors that explain the observed inequality in labor markets are overpopulation and colonial legacy. Therefore, there is no reason to reject the model, and we may accept sigma theory at this stage of our research.

Up to now, short-run static models of the epsilon, omega, and sigma theories have been presented. The empirical predictions derived from each model have not been refuted by the relevant empirical regularities of the First World and Third World countries listed in chapter 2. Facts 1 to 4 have thus been explained by these partial theories of capitalism. In order

to confront these theories with Facts 5, 6, and 7, dynamic models showing endogenous capital accumulation (physical and human capital) are needed. In the process of constructing dynamic models, the social cost of inequality in the functioning of capitalist societies must be understood. This is the theme of the next chapter.

CHAPTER 7

Inequality and Social Disorder

According to the epsilon, omega, and sigma models presented so far, production and distribution in capitalist societies are endogenously determined. Income inequality is thus an outcome of the economic process. An implicit assumption of these models was that the degree of inequality was always socially tolerable. Therefore, once static general equilibrium was reached, production and distribution could be repeated period after period under the same rules of the economic game, that is, the general equilibrium implied social order.

This prediction is however inconsistent with the most distinctive features of the workings of capitalism. Workers go on strike; people expropriate assets or incomes of others by force; cities are insecure places to live in; democracy is interrupted by dictatorship; and so on. Governments spend significant resources in protecting private property and individual security through police, courts, and jails. People also make significant expenditures to protect their property directly, which gives rise to the development of the security industry. These are reflections of social disorder.

Social disorder will be defined here as the behavior of people directed to break the rules of the institutional context. What are the factors that lie behind the observed social disorder? General equilibrium with excess labor supply is a trait of capitalism, as shown in previous chapters. Is general equilibrium with social disorder another trait of capitalism?

This chapter presents, first, a theory that intends to explain the role of inequality in the generation of social disorder in the three types of capitalist societies; second, it also presents a theory of government behavior toward inequality. Then both theories are integrated into new general equilibrium

models of the epsilon, omega, and sigma theories. The empirical predictions of these models are derived and then confronted against facts.

Limited Tolerance for Inequality

In order to understand the origin of social disorder, a general theory will be presented in this section. We may call it the *theory of limited tolerance for inequality*. The following alpha proposition is then proposed:

> α*(C). (1) Limited tolerance for inequality*: In capitalist societies, which are unequal societies, individuals have a sense of justice or fairness with respect to economic inequality, which sets thresholds of tolerance for inequality. Whenever inequality reaches levels beyond their tolerance thresholds, individuals will react and seek measures (legal or illegal) to restore inequality to the tolerable range.

The symbol "C" after alpha indicates that this theory is "general," in the sense that it is intended to be valid in the three capitalist societies that have been studied, namely, epsilon, omega, and sigma. Indeed, limited tolerance for inequality is included as an assumption of their alpha propositions, which will now be developed.

As mentioned before, capitalism is an unequal society; moreover, in this society, people act motivated by self-interest. In this context, people will also seek social status, which may then be treated as a good. The individual's relative position in society in terms of real income is thus a very important objective. Inequality becomes an essential factor of human behavior, as the proposed theory says. Individuals will seek not to be left behind; a fall in real income will not be tolerated. Individuals will seek to earn higher incomes in order to gain social status or maintain it and "keep up with the Joneses." If general equilibrium with equal incomes for all was attained, people would hardly seek higher incomes, nor would they worry about being left behind. However, in an unequal society, people act motivated to seek higher relative incomes.

The theory of limited tolerance for inequality assumes the existence of a threshold level of tolerance for inequality. Whenever inequality reaches values that go beyond the threshold, individuals will start considering inequality to be unjust or unfair. In addition, the theory assumes that people do not just contemplate injustice, but react, do something, protest, and seek to resist what they consider to be unfair distribution, or else retaliate with illegal actions. The social conflict of income distribution becomes important in society because of the existence of the tolerance threshold; if such a

threshold did not exist, any degree of inequality would be socially tolerable and thus extreme inequality would not generate social disorder.

Following Popperian epistemology, the first criterion to accept this theory is that it must be empirically falsifiable. For this purpose, a particular model of the limited tolerance theory is now developed. The following auxiliary assumptions will then be introduced. The model will assume that individuals will not tolerate a fall in the current real income. The threshold value of the relative income (compared to the income of the relevant social group) is subjective and thus unobservable, which means it is an *exogenous element*, not an exogenous variable; so, it must be assumed to be fixed if the model is going to be falsifiable.

The set of income distributions that is socially tolerated will be called the *region of social tolerance for inequality*. This is a strict subset of all possible distributions of income, that is, the region consists of various distributions, not a unique distribution.

Figure 7.1 illustrates the workings of the model. Suppose society is composed of two social groups A and B, where B is richer. Their current incomes and their thresholds of relative incomes determine their limits of tolerance. The region of social tolerance is thus given by the shaded area D*. Perfect equality (point E) does not belong to the region, as the wealthy

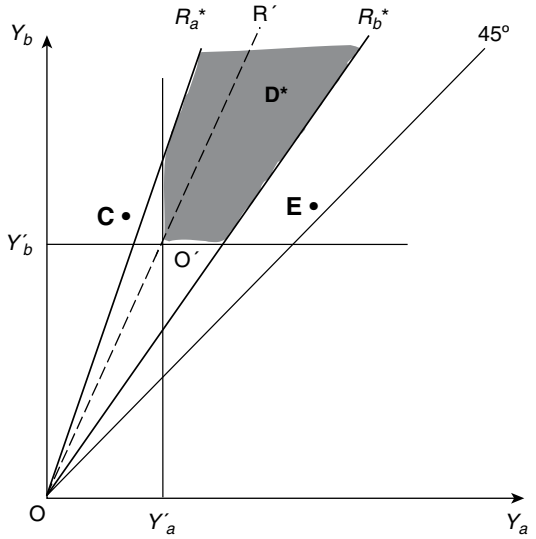


Figure 7.1 Region of social tolerance for inequality.

(group B) will not tolerate this situation. Group A will not tolerate situation C. Figure 7.1 assumes that tolerance for inequality is determined in terms of proportional differences in income. (We could also assume tolerance in terms of absolute differences in income, which would define D* by parallel lines to the 45° line.)

Any income distribution that lies outside the social tolerance region will lead to reactions from one of the social groups, which will include breaking institutional rules, for the group will consider the excessive inequality the result of an unfair institutional framework. Markets and the rules of democracy, the basic institutions of capitalism, will thus be challenged. These actions will lead to social disorder.

The region of social tolerance does not imply that social groups are indifferent to every possible income distribution within the region. There is still an area of social conflict about income inequality, in the sense that social groups would prefer a higher relative income; but income inequality that lies *within* the region D* does not lead to social disorder.

There is nothing in the general equilibrium models presented so far that ensures that income inequality will fall within the region of tolerance. It is not a necessity. Thus, this region cannot be called "equilibrium inequality" because there is no mechanism by which income inequality that is outside the region of social tolerance can return to the region spontaneously. In contrast to a Walrasian market, such as the potato market, in which market equilibrium price is self-regulated, tolerable income distribution is not self-regulated by market forces. Democracy has no mechanisms to restore equilibrium spontaneously either, as will be shown next.

Now assume that individuals have different threshold values of tolerance for inequality. For a given degree of income inequality, there will be a group of people that tolerates it and another that does not. If the degree of inequality increases further, the group that did not tolerate the initial income inequality will certainly not tolerate this higher inequality; but from the group that did tolerate the initial income inequality, there will be a subgroup that will not tolerate this higher inequality now. Therefore, there will be more people that do not tolerate the higher degree of inequality.

The empirical prediction of the limited tolerance model is that in the aggregate there will exist a positive relation between income inequality and social disorder. The empirical prediction of the model between social disorder (SD) and the degree of income inequality (D) can then be written as follows:

$$SD = f(D/D^*),\ f' > 0,\ D > D^* \tag{7.1}$$

Now social disorder will appear when a significant proportion of the population considers the current degree of inequality (D) too high with respect to the tolerable value (D^*); moreover, social disorder will increase as inequality rises further beyond that threshold value.

Income inequality is an outcome, an endogenous variable, of the economic process. As shown in previous chapters, in each type of capitalist society, the exogenous variable that determines the *level* of income inequality is the initial inequality in the individual distribution of economic and political assets (δ), whereas the other exogenous variables will cause small variations around this level and may thus safely be ignored, that is, the initial inequality is the essential factor in the determination of income inequality. Then equation (7.1) can be written as

$$D = g(\delta),\ g' > 0 \tag{7.2}$$

$$SD = F(\delta/\delta^*),\ F' > 0,\ \delta > \delta^* \tag{7.3}$$

Social disorder now depends upon the values of the initial inequality, the excess value over what is socially tolerable.

A comparison of this theoretical model with other theories on inequality tolerance is in order. First, consider the theory of justice developed by the philosopher John Rawls (1971). His normative theory can be transformed into a positive economic theory as follows. Consider a society in which individuals face the same opportunities, in which everybody may go through the same risk of losses or the same windfall gains. In such a society, consider the following distributive rules:

1. If everybody is subject to the same risk of having no income, then people will agree on setting rules for a guaranteed minimum income for all, as a protection against that risk.
2. If everybody is subject to the possibility of extraordinary gains, then people will agree on setting redistribution rules so as to share part of that gain.

In such a context, any individual motivated by self-interest would independently design this distributive rule. It is an impersonal rule; it is symmetric; it is fair; thus, it is just. It will be acceptable as part of the social contract.

A community of fishermen composed of independent production units (where boats are privately owned) would accept these rules. If an individual ends the day with no catch, he will receive some fish from the rest. If an individual gets a very large catch, part of it will be redistributed. Because

everyone is subject to the same contingency, the distribution rules—actually the redistribution rules—lead to the region of social tolerance for inequality.

However, Rawls's theory implies a region of social tolerance that is symmetric with respect to the 45º line in figure 7.1. Hence, an empirical prediction of Rawls's theory is that perfect equality would belong to the region of tolerance, that is, perfect equality in society will lead to social order. This is a different prediction from what emerges from the theory of limited tolerance for inequality, in which the wealthy would not tolerate a new perfect equality situation.

Implicitly, the abstract world of Rawls is one in which there are no social classes. In this society, the probability of reaching any position in the distribution of income is the same for everyone; people play the economic game with unloaded dice. This is his assumption of people acting behind a veil of ignorance about the future. It is a rule for very flexible and open societies, where there are no institutional barriers that limit social mobility. In a society of class and citizenship differences, this could hardly be the case. The opportunities for earning incomes are different for different social classes. Rawls ignores the role of the initial inequality in the individual endowments of economic and political assets in income inequality determination.

Another important author on this topic of inequality tolerance is the economist Albert Hirschman (1973). He proposed the theory that any inequality is tolerated if it is expected to be temporal. In the initial stages of economic growth, when inequality is likely to increase, society's tolerance for such disparities will be substantial because people's expectations are that those disparities will be temporary. If those expectations are not met, then there will be social disorder. Hirschman uses as the metaphor of a tunnel in which there is traffic blockage. When cars start moving in one of the lanes only, people in the other lane will accept this disparity because they expect that shortly they will be moving as well. If this does not occur, social protest will emerge. Hirschman's idea is complementary to the limited tolerance theory, for it adds an underlying mechanism by which social disorder results from income inequality increases.

Government Behavior towards Inequality and Social Disorder

A question now arises: What would be government behavior toward inequality and social disorder?

Social order is certainly a public good. It is a good that is nonrival and nonexcludable. The first characteristic implies that the consumption of

social order by an individual does not deprive others from consuming it; the second, says that nobody can be excluded from consuming social order once it is produced.

The provision of public goods usually falls in the domain of the state because of the lack of market incentives to produce them. Firms have the incentive to produce only goods that are privately consumed, which can then be sold in the market, and which can thus generate profits. (We cannot buy some units of social order together with some kilos of potatoes in a supermarket.)

Public goods play an important role in society. The presence of public goods helps to construct a human society with social ties and common interests; if it were not for public goods, a capitalist society would just be a sum of individuals, each with personal interest. Because market relations are impersonal in nature, social ties are not essential. The production of public goods leads society to attain the common good. In capitalist societies, governments are in charge of the provision of public goods, such as money supply, infrastructure capital, and also social order.

Standard economics usually assumes that government behavior is exogenously determined in the economic process. Governments are even told what to do. Consider the fact that almost every economic study concludes with a section or chapter about policy recommendations, in which the study tells governments what they should do in order to solve a particular problem. This literature assumes, implicitly, that governments do not pursue interests of their own.

Another view within standard economics comes from public choice theory. The assumption here is that people when in government act motivated by self-interest, as in any other activity. Thus, politicians exhibit selfish behavior. They may also have other motivations, political or ideological, but these are not considered the essential factors underlying their behavior.

Political power originates in the particular form that democracy adopts: representative democracy. Participatory or direct democracy would not generate such power. Most democratic systems in the real world operate under representative democracy, elected through universal voting rights. Under such a system, we may assume that politicians seek political power and incomes that go with running the state apparatus; thus, their social function of seeking the common good might be accomplished, but only as a by-product of their selfish motivation, as some authors have pointed out (cf. Downs 1957; Orchard & Stretton 1997; Persson & Tabellini 2000).

The theory of government behavior that will be adopted in this book will assume the form of representative democracy wherein politicians act motivated by self-interest. The following assumption is then proposed:

α(C)(2). Politicians act motivated by self-interest: In a capitalist society, politicians seek two objectives hierarchically ordered: first, political power and only then income, subject to pressure group demands, the fiscal budget constraint, and institutional constraints.

This theory assumes the existence of a political class. As in the case of capitalists, politicians first seek to remain in the privileged position of membership of a social class: the political class. The objective of income maximization is subject to the attainment of the first objective, not as a substitute for it.

According to this theory, governments act out of selfish motivations; they are social actors, as capitalists or workers, for they interact with other social actors, that is, they are not above social classes. Given the values of the exogenous variables in the economic process, governments will choose the values of the endogenous variables according to their rationality. Changes in the exogenous variables will modify their choices.

To put this theory of government behavior into the falsification process, auxiliary assumptions are now introduced to construct a particular model of the theory. First, governments will seek to maximize votes (or popularity, which is a form of voting) subject to the public budget constraint, institutional constraints, and the constraints arising out of the demands of the pressure groups. The structure of the public budget constitutes the endogenous variables. Second, votes will depend positively on the performance of the aggregate variables of the economy, such as output, employment, income inequality, poverty rates, and inflation rates. The objective of maximization of votes is pursued through mechanisms of shifting preferences, which include visible outcomes of public policies, as pointed out earlier, and also campaigns and propaganda. In seeking votes maximization, politicians will thus require financing, and evaluation of expected rates of return, as is the case in any investment that people make.

Third, democracy is usually defined as the "government of the people, by the people, and for the people." Nevertheless, representative democracy is a way to concentrate political power in governments. Why do people accept this type of democracy? Under capitalism, where people act motivated by self-interest, voters have the incentive to act as free riders on public matters: "Everybody's business is nobody's business." Governments are confronted with a dilemma about accountability. Governments should be accountable not only to the workers, who constitute the mass of voters, but also to the capitalist class, for they own the capital. Given that votes can be bought by other means, governments will be more accountable to the capitalist class than to the workers. The democratic system will thus tend to be plutocratic.

According to this model, money can buy everything in the capitalist society, including votes.

The next logical step is to determine the equilibrium conditions in government behavior. Given the exogenous variables, governments will seek to reach their best positions, their equilibrium positions, with the objective of maximization of votes. The following forms of government behavior may then be derived from the model:

1. Governments will seek to allocate the public budget to discretionary expenditures (with which to buy votes) rather than to mandatory expenditures (on securing citizens' access to their rights, which do not buy votes).
2. Governments will seek to finance fiscal expenditure with more debt rather than more taxes, as the latter will affect negatively next elections, whereas the former will not.
3. Governments will seek to use fiscal policy according to the political cycle, that is, governments will seek to increase public expenditure before elections and reduce it after elections. Fiscal policy is dependent on the political cycle.
4. Governments will seek to allocate discretionary public expenditure by industrial sectors and regions according to their political profitability, that is, expenditure will be higher in the most densely populated areas (urban rather than rural areas) and in the most visible projects (school buildings, which can be inaugurated, rather than education reform to increase its quality, which cannot be inaugurated).
5. Public expenditure that creates dependency (as clients) will be preferable to those that do not. Cash transfers to the poor that keep them poor and always in need of transfers will be preferable to, say, investment in schooling that gives them more economic autonomy.

These propositions just refer to the equilibrium situations, given the values of the exogenous variables. Because these equilibrium conditions are observable, they constitute empirical predictions of the model as well. The model thus predicts governments to exhibit myopic behavior: they will seek to give priority to public policies that have short-term effects upon gaining preferences for the next election rather than upon those having long-term effects.

These empirical predictions of the model have rarely been studied in the international empirical literature. Prediction (3) was corroborated in a study about the US government behavior, in which it was shown that government social expenditure indeed increases before elections and falls after elections (Rogoff 1990).

How does the government behave toward inequality? American economist Arthur Okun (1975) pointed out long ago that democratic capitalism operates with a double standard: the political system preaches equality, but the market system operates with inequality. In order to make the system viable, Okun proposed the theory that society seeks to establish some rules to set limits to inequality and thus ensure social order. The state establishes rights aimed at this objective. Some goods and services are taken out of the marketplace and distributed to the population as public goods. Money therefore cannot buy everything in society. One could consider this an institutional rule for the functioning of society. This rule would be part of the social contract.

The question is whether social actors have the required power and incentives to establish this type of social contract. Where do the political demands for rights come from? Okun does not solve this question. Implicit in Okun's theory is that the society he is referring to is a society in which people are citizens of the same class (epsilon society). The situation would be different in a capitalist society in which citizens are unequal, first class and second class.

Assume, as before, the existence of two types of abstract capitalist societies, called epsilon and sigma. (Ignore omega for the time being.) They differ in the two initial conditions: factor endowments and initial inequality. Epsilon is more endowed with capital per worker than sigma; also initial inequality is lower in epsilon than in sigma, where initial inequality refers to the distribution of economic assets and political entitlements among individuals (δ); in particular, epsilon is a class society (with unequal distribution of economic assets), but here political entitlements are equally distributed, whereas sigma, which is also a class society, has political entitlements unequally distributed, so that there are first-class and second-class citizens. Suppose the concentration of capital is similar in both societies. Then, the initial inequality is higher in sigma than in epsilon ($\delta_\sigma > \delta_\varepsilon$). The idea is that these two abstract societies will resemble well the basic structure of the real-world capitalist system: the rich and poor countries, named here the First World (epsilon) and the Third World (sigma).

Now government behavior will be different in each type of society. The government will respond to the demands of the citizens for rights in epsilon society, but not in sigma; hence, more rights will be granted in the first compared to the second. The government model therefore predicts that mandatory expenditure (allocated to the financing of rights in the form of public goods) as a proportion of total government budget will be higher in epsilon than in sigma; moreover, the provision of public goods as part of total output will also be higher in the former. The degree of social order will thus tend to be higher in epsilon than in sigma.

This prediction is consistent with the fact we observe in the First World and the Third World: government expenditure (and taxes) as a proportion of the GDP is higher in the former than in the latter. The citizenship effect is thus included in the income effect in this relation.

Given the set of rights in society, suppose income inequality still falls outside the region of tolerance. What will governments do in order to ensure social order in the short run? Governments could choose to redistribute income until inequality is placed in the tolerance region. The mechanism for this is public choice made democratically, which implies one person one vote (different from the market choice, which is one dollar one vote); moreover, a democratic decision is based on the majority rule, which implies that the choice of the median voter (which is placed at the center of all positions) will decide. Hence, in the case of income redistribution policies, the choice of the median voter will decide.

But does the median voter belong to the poor or the rich group? If the voter belongs to the poor group, then the redistribution policy will win. One may thus expect that in a democratic capitalism the majority rule, or the median voter rule, would endogenously move inequality to the socially tolerable region.

If the distribution of income was a normal distribution—symmetric distribution or bell-shaped distribution—the mean income would divide the population into two equal parts (50 percent below the average and 50 percent above the average). However, we know that income distribution in the real world is not symmetrical, but asymmetrical: the mean income will divide the population into two unequal parts (say 70 percent below the average and 30 percent above) because income concentration implies that for each rich individual (with income above the average) there will be many poor individuals. Therefore, the median income (dividing the population 50 percent below and 50 percent above) will be below the mean income, which implies that the median voter will have an income that is smaller than the average income and thus will belong to the poor group.

Democracy should therefore produce redistribution policies endogenously for income distribution to become a self-regulated process, not by the market system, but by the democratic system. If inequality rises, public policies chosen democratically will bring it back to the socially tolerable region and social disorder will hardly have any significance. However, this is not what we observe in the real world, neither in the First World nor in the Third World.

Why does the median voter democratic principle fail? There are several theoretical reasons. First, under representative democracy, politicians are given power. Therefore, voters do not choose policies, their representatives

do; moreover, the representatives are not a random sample of their voters, as doctors are not of their patients (Banfield 1958). Second, voters and representatives face the *principal-agent problem*. This problem appears whenever the objectives of the agent (the government in this case) are inconsistent with those of the principal (the voters). Third, the constraints put by the wealthy, a powerful pressure group, will go against income redistribution. As was assumed above, capitalist democracy is distorted by the use of money in the election process and in the public choice process (money after all buys everything!); hence, representative democracy operates as a plutocracy: it is the government by the wealthy and for the wealthy (Fukuyama 2011).

In contrast to epsilon society, sigma society democracy operates with a significant proportion of voters that are second-class citizens. They constitute the social group in need of rights, but it has no voice to demand them precisely because it is excluded from basic rights. This group lacks the right to have rights. Therefore, governments in sigma society will have even much less incentives to be accountable to the masses and more incentives to rely on the more direct mechanisms to buy votes. Governments may then have more incentives to use a policy mix that contains more repression than redistribution. Repression measures imply stronger law enforcement or enacting new laws or both, or even disrespect for rights. While repression has immediate results to restore social order, redistribution will show results over longer periods, maybe after the next elections.

In sum, the political behavior model predicts that governments will behave differently in the two types of capitalist societies; although governments seek the same objectives, they will face different constraints. Compared to epsilon society, in sigma society there are fewer rules (rights) to set limits to inequality; hence, discretional expenditure is more predominant in the fiscal budget. Discretional government expenditure is precisely the instrument to buy votes and it is relatively more significant in sigma society. In both societies, income distribution is not a self-regulated process. Excess income inequality cannot be reduced endogenously by the market system or by the democratic system even if it leads to social disorder. Governments have no incentives to reduce the excess income inequality and place it within the socially tolerable region, for that would imply redistribution measures and confrontation with the economic elites.

General Equilibrium with Excess Inequality and Social Disorder

Can general equilibrium with an excessive degree of inequality and social disorder exist?

Consider the general equilibrium models of epsilon, omega, and sigma societies shown in previous chapters. The prediction of these models is general equilibrium with an excess of labor supply. Suppose the market equilibrium of prices and quantities leads to income inequality that lies outside the tolerance region. The prediction of the model is that social disorder will be significant. Social disorder will negatively affect the production process. The mechanisms are now explained.

Under social order, total output and the degree of inequality that result from the economic process would be repeated period after period as long as the exogenous variables remain fixed. Under social disorder, in contrast, the production and distribution process becomes stochastic. The production process cannot be repeated period after period even if the exogenous variables remain fixed. Some losses in the flow of output or in the stock of capital will occur due to social disorder. Social disorder affects the normal production process through periods of interruption due to workers' strikes, social unrest, property insecurity, and political instability. These shocks reduce the flow of output compared to that of periods with social order.

Social disorder also implies transgression to property rights, weak respect for the rule of private property rights. Workers will take actions to reduce inequality through forced (illegal) private redistribution of income or assets. This redistribution may be considered as equivalent to a lump sum tax levied upon the capitalists. Production is now subject to the risk of external shocks upon the total output and physical capital. Production and distribution will then take different values, depending on the occurrence of shocks; therefore, it is the mean value or the expected value of both total output and the degree of inequality, together with a given variance, that will be repeated period after period as long as the exogenous variables remain fixed.

Figure 7.2 represents general equilibrium under social disorder in epsilon society. Just to simplify, consider only two possible events: shock and no shock, with probabilities P_1 and P_2. The aggregate marginal productivity of labor is then represented by two curves: $A'M_1$ with probability P_1 (event of shock) and DM_2 with probability P_2 (event of no shock). The expected marginal productivity of labor is given by curve BM^e, which is just the weighted average of both curves, where the weights are given by the probability values. The superior curve DM_2 constitutes the productivity frontier of society. Taking the expected marginal productivity curve BM^e as given, and also the market real wage rate (w) as given, labor market equilibrium would be at point E, where firms would hire OH workers. Expected total output would be equal to the area under the segment BE. Expected profits would be equal to area wBE and total wage bill would be the area OwEH.

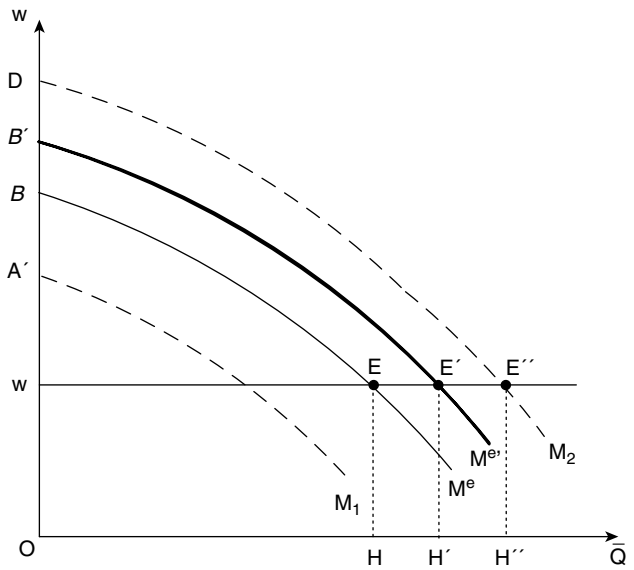


Figure 7.2 General equilibrium with social disorder in epsilon society.

Figure 7.2 may also represent the capitalist sector of the sigma society. Given that the subsistence sector is residual, the conclusions obtained in the epsilon model will also apply to the sigma model. Therefore, a *generalized epsilon-sigma model* will be developed in what follows, just to simplify the presentation. This model will also be useful to establish the differences between these two societies. It also follows that this model will apply to the omega society as well.

Under an excessive inequality and social disorder context, the production process will be subject to the risk of private redistributive shocks, which will determine the mean value of total output together with the risk or variance. If firms and the government took social disorder as exogenous, then the solution shown in figure 7.2, at point E, would constitute the equilibrium situation.

Nevertheless, firms and governments will react to the initial social disorder. Under the context of social disorder, in which shocks of forced private redistribution of income and assets take place, what will capitalists do? Because capitalists in this context will seek to maximize the value of the expected profit, subject to risk levels that are bearable, they will react to the situation by acquiring more inputs to protect their property rights. Some quantities of inputs destined to protect private property

(such as fences and security systems) will be added to the production process, even though they are not necessary from the technological point of view. In addition to *productive employment*, firms will hire more workers as *protective employment*. Assume that the protective measures taken by firms indeed reduce the risk of shocks on the production process. It is as if firms were seeking to be insured (although not fully) with the protective measures.

Because capitalists seek to maximize expected profits, firms will be willing to increase their production costs in the form of protective costs if the resulting total *net* expected profits (net of the cost of protection) would be higher compared to doing nothing. Indeed, they will increase these costs until it becomes equal to the increase in expected profits. This optimum total cost in protective inputs will be reached by firms, which will then become a fixed cost. To this fixed cost there will then correspond a higher expected marginal productivity of labor curve.

On government behavior, it will be directed to a combination of repression and redistribution, which will also reinforce the fall in the probability of shocks. In the short run, repression may have priority so as to reduce the degree of social disorder for the coming elections. As a result of the reactions of firms and the government, workers will in turn react by reducing the activities directed to forced redistribution in view of the increasing costs (penalties) for this behavior.

Assume that the interactions among social actors converge to a general equilibrium situation, in which no social actor will have the power or the incentive to change it. There will exist a nonzero probability of forced redistribution shocks that is of equilibrium. At this probability value, capitalists do not wish to increase their protection costs and the government does not wish (or is unable) to apply more repression or more redistribution. The solution implies nonzero probability of the occurrence of redistributive shocks because the economic and political elites have no incentive to make the adjustments needed to reduce income inequality to the level of social tolerance. It is a result of the lack of incentives to produce the public good social order. If capitalists and the government had had these incentives, they would just have acted redistributing income; then protective costs and repression would have been unnecessary.

The equilibrium value of total output under social disorder in epsilon society can be written as follows:

$$\bar{Q} = \gamma(\delta) f(K, L) = F(K, L, \delta), \gamma \leq 1, \gamma'(\delta) < 0, F_3 < 0, F_i > 0 \tag{7.4}$$

$$\bar{Q}/L = \bar{q} = \gamma(\delta) g(k) = G(k, \delta), G_1 > 0, G_2 < 0 \tag{7.5}$$

Equation (7.4) indicates the assumption that, given the stock of capital (K) and the quantity of hired workers (L), firms will produce different values of expected total output ($\bar{Q}$) depending on the degree of social disorder, which in turn depends upon the degree of the initial inequality (δ). The productivity factor γ depends upon the initial inequality and can have a value $\gamma = 1$ when the initial inequality ($\delta \leq \delta^*$) is conducive to degrees of income inequality that fall within the tolerable region, which implies social order. It will be less than one ($\gamma < 1$) when the production process takes place in a context of social disorder ($\delta > \delta^*$). Moreover, the higher the degree of initial inequality, the lower the productivity factor γ and thus the lower the expected total output. The mechanism is that the higher the initial inequality, the higher the income inequality, and therefore the higher the social disorder.

Equation (7.5) shows that the expected average productivity of labor, which is derived from the previous equation under the assumption of constant returns to scale technology, depends negatively upon the initial inequality and positively upon the capital labor ratio ($k=K/L$), that is, for a given capital-labor ratio, output per worker will depend negatively upon the initial inequality. The expected marginal productivity of labor will have the same determinants and with the same effects.

Figure 7.2 shows general equilibrium with social disorder at point E′. The adjustments made by the social actors to the initial situation of social disorder imply a fall in the value of the probability of shocks P_I, which in turn implies a shift of the expected marginal productivity of labor from curve BM^e to curve $B'M^{e\prime}$. Given the same market real wage rate, the equilibrium level of employment will be OH′, which leaves a quantity of excess labor supplied. Expected total output is equal to the area under the segment B′E′. The area OwE′H′ indicates the wage bill and the area wB′E′ the net expected profits (expected profits minus protection costs).

The labor and commodity markets constitute the core of general equilibrium, as mentioned earlier. Figure 7.2 shows that the labor market is in equilibrium (with excess labor supply). It also shows that the commodity market is in equilibrium: the quantity demanded is equal to the wage bill and the quantity supplied is equal to total output minus profits, which is just the same area of the wage bill. The implication is that the money market is also in equilibrium, which just follows from equilibrium in the commodity market: workers and capitalists willingly hold the quantity supplied of money; for if this were not the case, part of the real income would be utilized to increase or decrease cash balances. Therefore, we have attained general equilibrium with excess inequality and social disorder in epsilon society.

Regarding sigma society, equations (7.3) and (7.4) will refer to its capitalist sector only. However, general equilibrium in sigma implies that the size of the subsistence sector is residual, that is, it is determined once employment in the capitalist sector is known. Therefore, the conclusions obtained for epsilon also apply to sigma. Thus, the solution shown in figure 7.2 is general, as it applies to both types of capitalist societies. Of course, the function $\gamma(\delta)$ will be particular to each type of society, so will the value of δ. In sum, general equilibrium with excess inequality and social disorder is a trait of every type of capitalist society.

General equilibrium with social disorder has several features. Given the exogenous variables, production and distribution of equilibrium are determined. In this type of equilibrium, the social environment is one of fear, distrust, and risk due to the shocks of private redistribution that capitalists and workers will confront. The equilibrium is static and stochastic: the mean values and the variance (degree of risk) of the endogenous variables will be repeated period after period as long as the values of the exogenous variables remain fixed. Equilibrium with social disorder implies high transaction costs in the functioning of society.

Another feature is that general equilibrium implies low productivity relative to the productivity frontier of society. Excessive inequality thus generates inefficiency in the economic process. The model also predicts that *social order* is an important public good that affects the economic process, just like other conventional public goods, such as social infrastructure (bridges and roads), or like natural disaster shocks.

In light of this general equilibrium model, some distinctions need to be made about the economic process. The income distribution that emerges from the market and democratic institutional rules may be called the *primary distribution*. In this case, fiscal policy is introduced in the economic process, with tax rates and transfer rates as new exogenous variables. Total tax revenue, which will be equal to the total government expenditure on public goods, and the distribution of public goods (as universal and local) among social groups will be endogenous. Income inequality will include the net government transfer (net of taxes) to each social group. Note that no separation is analytically justified between distribution from the market and distribution after government net transfer. Once the tax and transfer rates are fixed, people will seek to translate the tax burden to others and press to get more transfers; hence, general equilibrium will reflect the result of those interactions.

If the primary distribution falls within the social tolerance region, it will be the final income distribution outcome; if it does not, and as a response to the generated social disorder, government behavior may seek to redistribute

income by changing the net tax rates and transfers. A new general equilibrium will emerge with a new income distribution. This may be called the *secondary distribution*. If the new inequality falls within the tolerance region, it will be the final income distribution; if it does not, private forced redistribution will take place and intend to modify the secondary distribution. Thus, this may be called the *tertiary distribution*, and that will be the final static general equilibrium with the latent risk of forced redistribution shocks, that is, general equilibrium with social disorder.

In sum, the epsilon-sigma models are able to predict general equilibrium with social disorder. Compared to general equilibrium with social order, general equilibrium with social disorder has significant features. First, it is inefficient, for society does not operate on its productivity frontier. Second, society operates as a high-risk society, subject to shocks during the process of production and distribution. High-risk society implies a lower quality of life for the people living in it. Third, equilibrium with social disorder is a second-best solution in the functioning of capitalism. Forced redistribution of income does not challenge the capitalist system; it is just a way to make the very unequal society workable, but with high transaction costs. Fourth, the nature of general equilibrium with social disorder or excess inequality points out the existence of another disease of capitalism, similar to that of general equilibrium with excess labor supply.

Beta Propositions

So far the theory of limited tolerance for inequality and the theory of government behavior have been introduced into a general equilibrium model of epsilon and sigma societies in which the initial inequality is too high ($\delta > \delta^*$). The general equilibrium situations have then been established. The conclusion is that general equilibrium with social disorder is an outcome of the economic process, which is still static, but stochastic; thus, the *expected or mean values* of the endogenous variables will be repeated period after period as long as the exogenous variable initial inequality remains unchanged. The other exogenous variables (factor endowments and money supply) will alter equilibrium income inequality but around the *level* determined by the initial inequality. This equilibrium result is valid for each type of capitalist society.

Consider now the effect of changes in the initial inequality (δ) on the general equilibrium with social disorder. The general equilibrium is clearly stable and comparative statics applies. According to the epsilon-sigma model, income inequality underlies social disorder and initial inequality underlies income inequality.

Figure 7.3 shows the general equilibrium relationships between inequality and social disorder in any type of capitalist society. The vertical axis measures the degree of income inequality. The horizontal line has two axes: to the right of the origin O, the degree of inequality in the distribution of economic and political assets is measured; and to the left, the degree of social disorder. Hence, the graph has two panels. Panel (a) shows the relationship between asset inequality (δ) and income inequality (D), in which the curve MN represents the secondary income distribution. The initial assets inequality determines the *level* of income inequality; however, small variations around this level will take place due to changes in other exogenous variables. Panel (b) shows the relationship between income inequality and social disorder. The threshold value is marked with D*, which corresponds to δ*.

Below these threshold values, income inequality can take the values, which are socially tolerable, along segment ME, and the corresponding degree of social disorder lies along the curve AB. For example, initial

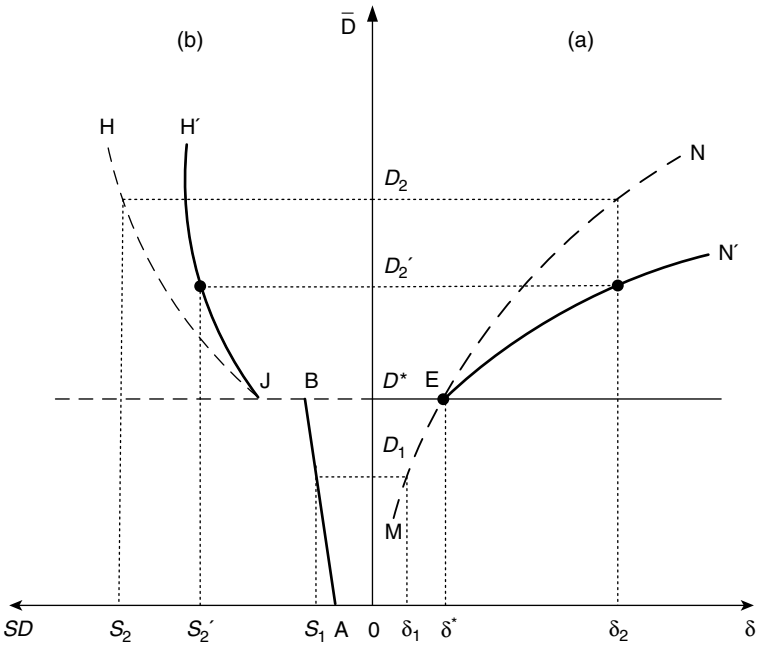


Figure 7.3 Relationships between asset inequality (δ), income inequality (D), and social disorder (SD).

inequality δ_1 determines income inequality D_1, which implies a small degree of social disorder S_1. These are general equilibrium values.

Above the threshold values, such as δ_2, income inequality would initially take the value D_2, along the segment EN, an extension of segment ME, whereas the social disorder would jump to curve JH and take the value of S_2. The final general equilibrium situation will however take the value D_2', along the segment EN′, which reflects the adjustments made due to the mechanisms of forced redistribution and thus represents the tertiary distribution of income; finally, the corresponding social disorder values will take the value S_2', along the new curve JH′.

Whenever the degree of inequality takes a value beyond the threshold value, there is some degree of adjustment to reduce it, but the model assumes that it is only a partial adjustment, not a total one, that is, there is no spontaneous mechanism by which the excessive inequality is moved to a level below the threshold value, to the region of social tolerance. According to the epsilon-sigma model, income distribution is not a self-regulated process; it can be regulated neither by the market nor by democracy, the two basic institutions of capitalism. Thus, for a sufficiently high initial inequality (δ), general equilibrium with a high degree of social disorder, corresponding to the high degree of income inequality will be the outcome of the economic process in any type of capitalist society.

It should be noted that equations (7.1) and (7.2) are represented in figure 7.3. The relationships presented there now correspond to the general equilibrium with excess inequality. The reduced form—equation (7.3)—can also be visualized in figure 7.3.

Up to now, the models have been very general, applicable to any of the two types of capitalist societies. Hence, the question now is this: What does the model predict about differences between them?

The initial inequality in sigma is higher than in epsilon, as indicated above. One reason for this is the assumption of the existence of first-class and second-class citizens in sigma and a single class in epsilon. We may now include the assumptions that the concentration of capital ownership is similar in both societies, but the concentration of human capital is higher in sigma than in epsilon, due to the existence of different classes of citizens. Because the initial inequality is higher in sigma society, it follows that the level of income inequality must also be higher there.

The degree of social disorder does not depend upon income inequality alone, but also upon the society's threshold value of tolerance for inequality. Equation (7.3) and figure 7.3 will help us in understanding this difference in threshold values between the two societies. Consider, first, the assumption that the threshold values are the same in the two societies and that general

equilibrium with excess inequality applies. In this case, sigma society will show a higher degree of social disorder as it operates with a higher level of inequality.

An alternative assumption could be that thresholds of tolerance are subject to path dependence, that is, they are historically determined. The threshold value arises from the resistance of people to higher degrees of inequality to those experienced in the past. Therefore, the history of inequality is fundamental in the determination of today's social tolerance for inequality. Higher historical values of inequality will thus generate higher threshold values of social tolerance for inequality; hence, threshold values are different among societies. It is higher in sigma societies because citizenship inequality allows higher degree of tolerance for inequality.

A third factor in determining the degree of social disorder is the adjustment functions. In figure 7.3, let us consider segment EN′ to correspond to sigma society, and a lower curve (not drawn) to epsilon society. This would imply that the redistribution policy is much more significant in epsilon than in sigma (whereas repression is more significant in sigma); hence, social disorder will be lower in epsilon even holding constant differences in the degree of inequality and the threshold of inequality tolerance. Hence, social disorder also depends upon whether societies have created institutions to avoid socially intolerable income inequality outcomes. Epsilon societies have created rights to do that, but sigma societies have not. In epsilon, less social disorder would be needed to reduce income inequality because of the existence of economic rights to redistribute the primary income distribution. In sigma, by contrast, forced private redistribution and the consequent social disorder is the only mechanism that workers can use to redistribute the primary income inequality. Government behavior will depend upon the type of society: it will choose more repression and less redistribution to restore social order in sigma societies compared to epsilon.

In sum, the epsilon-sigma model predicts a higher degree of social disorder in sigma societies compared to that in epsilon societies. The essential factor explaining this result is the differences in their initial inequality.

Figure 7.4 displays the basic predictions of the model on the relationships between inequality and social disorder across the two types of capitalist societies. The horizontal axis measures the degree of income inequality and the vertical axis the degree of social disorder (SD). The curve AB corresponds to epsilon, whereas the curve EF corresponds to sigma. The difference comes from different values of the exogenous variable: the initial inequality (δ). Suppose both societies operate in the range of income inequality that is out of the region of social tolerance and thus with social disorder. Equilibrium in epsilon society is at point M and in sigma society at point N. The prediction

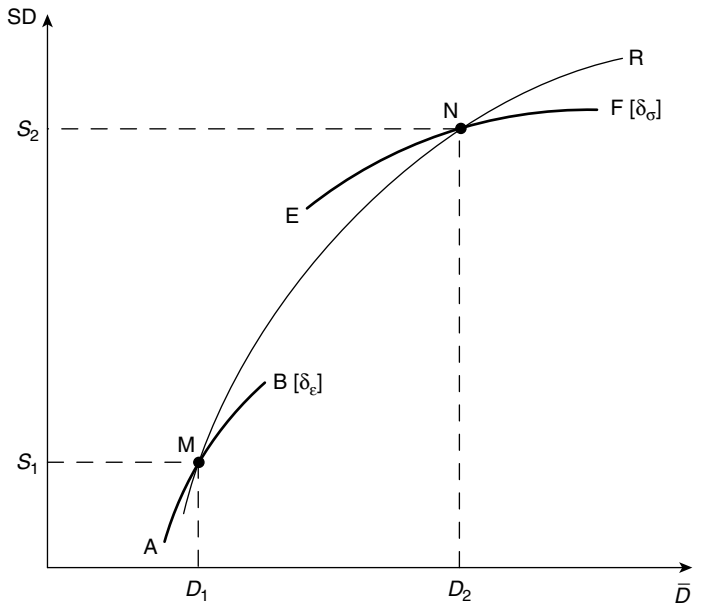


Figure 7.4 Income inequality (D) and social disorder (SD) levels by types of capitalist societies.

is that the epsilon operates in the region of low inequality and low social disorder; whereas sigma does in the region of high inequality and high social disorder. The relationship across epsilon and sigma societies is represented by the curve R, a curve that goes through points M and N. Omega society can now be introduced by assuming that it will lie somewhere between points M and N, possibly closer to N. This relationship is empirically observable and makes the model falsifiable.

It should be remarked that the prediction that more unequal societies show more social disorder leaves open the question of whether social disorder takes the form of individual or collective violence. The model does not necessarily predict collective violence because this would require collective action, which depends on other factors to materialize (cf. Olson 1965).

The reduced-form equations of the model for total output and distribution can now be presented. National income in epsilon society will be equal to the production in the capitalist sector, whereas in omega and sigma it will include the output of the subsistence sector. Hence, the expected or mean national income of equilibrium ($\bar{Y}°$) for each capitalist society can be represented as the following system of equations:

$$\bar{Y}^{\circ}_{j} = F^{j}\,(S_{mj},\, z^{*};\, \delta_{j},\, K_{j},\, S_{hj}),\, j = \varepsilon,\, \omega,\, \sigma \qquad (7.6)$$
$$F_1 \geq 0,\, F_2 > 0,\, F_3 \leq 0,\, F_4 > 0, \text{ and } F_5 \geq 0.$$

These are the reduced form equations of the model, one equation for each type of society, as indicated by the superscript *j*, which indicates not only differences in the function, but also in the range of values of the exogenous variable δ. In each type of society, factor endowments and the initial inequality will determine the *mean value of the level* of national income, with short-run variations around this level, which will be due to changes in the other exogenous variables, money supply and international terms of trade.

Similarly, the expected value or mean value of the degree of equilibrium income inequality for each capitalist society can be represented by the following system of equations:

$$\bar{D}^{\circ}_{j} = G^{j}\,(S_{mj},\, z^{*};\, \delta_{j},\, K_{j},\, S_{hj}),\, j = \varepsilon,\, \omega,\, \sigma \qquad (7.7)$$
$$G_3 > 0 \text{ and } G_i = (?) \text{ otherwise.}$$

These are also the reduced form equations of the model. In each society, the *mean value of the level* of the degree of income inequality is determined by the initial inequality, with short-run variations around this level that will be due to changes in the exogenous variables money supply and the international terms of trade.

These results can be compared to the static models of general equilibrium with social order, which were developed in chapters 4, 5, and 6. The reduced form equations for total output were presented in equations (4.16), (5.9), and (6.13), corresponding to epsilon, omega, and sigma, which are similar to system (7.13) above, with the exception that now the variable initial inequality has an effect (negative) upon total output, which is stochastic. The reduced form equations for the degree of income inequality were presented in equations (4.18), (5.11), and (6.13), corresponding to epsilon, omega, and sigma, which are all similar and equal to system (7.7), in which income inequality is a stochastic variable.

In sum, the exogenous variables that determine production and distribution in the short run are common to the three types of capitalist societies and include: factor endowments and initial inequality, together with money supply and terms of international trade. Beta propositions logically derived from the general equilibrium model for each type of capitalist society are given by the partial derivatives of the systems of equations (7.6) and (7.7). Actually, the causality relations are the same in the three types of societies and they can be utilized to empirically refute the theoretical models.

In qualitative terms, the general equilibrium models with excess inequality predict that capitalist societies operate in social disorder: the higher the former, the higher the latter. The exogenous variable is the initial inequality in the distribution of economic and political assets. Therefore, compared to epsilon society, sigma society tends to operate with higher social disorder, which implies weaker institutions, weaker law enforcement, and higher degree of corruption, that is, weaker democracy. Governments are less accountable due to the existence of first-class and second-class citizens and have no incentives to apply fiscal policies to redistribute income and make inequality more socially tolerable. The colonial legacy in sigma society maintains unchanged the *culture of inequality*. Thus, democracy operates differently in the Third World compared to the First World.

On the workings of market institution, the masses are not first-class buyers or sellers either; therefore, people have no incentives to comply with market contracts and norms. Thus, the culture of inequality is also embedded in the functioning of the market system, which operates with high transaction costs. Therefore, the market system qualitatively works differently in the Third World compared to the First World.

Empirical Consistency: The Capitalist World

A generalized epsilon-sigma model with results that apply to both types of societies has been constructed above. This model predicts that the Third World operates with a higher degree of social disorder relative to the First World. Social disorder is higher where income inequality is higher. Do facts refute this prediction?

The most important measures of social disorder found in the literature include property rights violations, political instability, and fragility of the rule of law. We should also recall that the theory of limited tolerance for inequality includes not only relative incomes but also absolute real incomes. Therefore, decline in absolute incomes of certain social groups would lead to social disorder independent of the change in relative incomes (inequality). For instance, periods of high inflation, in which real wages fall significantly, will be accompanied by social disorder.

Regarding property rights violations, an empirical study by the World Bank examined the statistical links between income inequality (measured by the Gini index from the data set constructed by Deininger &Squire [1996]) and violence (measured by crime rates from the UN *World Crime Surveys* data set). The sample includes 45 countries (16 from the First World and 29 from the Third World) for the homicide regression analysis, and 34 countries (14 from the First World and 20 from the Third World) for the

robbery regression analysis, for the period from 1965 to 1995. Both samples exclude countries from sub-Saharan Africa due to lack of data sets. The main conclusion of this study is that income inequality has a robust and significant positive statistical association with the incidence of both types of violence (Fajnzylber, Lederman, & Loayza 2002). Bourguignon (2000) found the same statistical association on a sample of 50 countries, using data of crime rates for the period 1985 to 1995 and the Gini index for 1985.

With regard to the prediction that higher income inequality leads to a higher degree of political instability, Alesina and Perotti (1996), in a sample of 70 countries from the First World and the Third World, covering the period 1960–1985, found a positive correlation between income inequality and sociopolitical instability. The 16 Latin American countries in the sample had the highest degrees of income inequality and also the highest degrees of sociopolitical instability.

Considering the prediction that higher income inequality leads to weaker democracies, political scientist Edward Muller, one of the most important scholars in his field, found a robust negative statistical relationship between the degree of inequality and the degree of democracy in a sample of 55 capitalist countries, which included 16 of the First World and 39 of the Third World, for the years 1960 and 1980 (Muller 1997). It seems clear that Third World countries show a higher degree of political instability than First World countries. The political history of Third World countries is one of experiences with socialist governments, dictatorships, limited democracies, and democracies, which meant interruptions of constitutions and drafting of new ones.

Regarding the rule of law effect, better rule of law indicators are found in the First World countries relative to the Third World, where the rule of law means that a country's formal rules are made publicly known and enforced in a predictable way through transparent mechanisms (World Bank 2001, Chapter 6, Figure 6.3).

In sum, empirical studies show findings that are consistent with the empirical predictions of the epsilon-sigma models of general equilibrium with social disorder. These studies are consistent with the relationship shown by the curve R in figure 7.4.

The standard literature also recognizes the existence of social disorder in the functioning of capitalism. However, it assumes, implicitly, that social disorder is exogenously determined, such that the rule of law can be improved by reforms in the judiciary system alone. In contrast, according to the epsilon-sigma model, social disorder is endogenous. Property rights violations, insecurity in cities, political instability, a weak democracy, and a weak judiciary system are all endogenous. What is exogenous is the

inequality in the distribution of economic assets and political entitlements, which generates excessive degrees of income inequality, and the consequent social disorder. As long as the power structure embedded in this inequality remains unchanged, capitalist societies will continue to operate with social disorder.

Unequal societies operate with inefficiency in the production process, as the potential productivity frontier of each society is not realized; another social cost is the low quality of life, for they become high-risk societies. Therefore, inequality is not only a normative or ethical question; it also plays a significant role in shaping the efficiency and the quality of a society. In most of the standard literature, inequality is seen as a normative problem only: excessive inequality is morally unacceptable. However, the theory of limited tolerance for inequality generates a scientific proposition, statistically testable: excessive inequality has significant social costs.

Conclusions

This chapter has presented particular models of the partial theories of the capitalist system—epsilon, omega, and sigma. They intend to explain the observed income inequality and social disorder in the First World and the Third World taken separately. These models include the assumptions that people have limited tolerance for inequality and governments act motivated by maximization of votes. Then, general equilibrium in each type of society exists with excess inequality and social disorder, which is a reflection of excess inequality in the distribution of economic and political assets. Markets and democracy, the two fundamental institutions of capitalism, have no mechanisms to avoid this feature of the general equilibrium.

Capitalist societies will thus continue to operate in this way as long as the initial inequality remains unchanged. The models predict that different types of capitalist societies function with different degrees of income inequality and social disorder, depending on their initial inequality: initial inequality, income inequality, and degree of social disorder are all higher in sigma society (the Third World) compared to epsilon (the First World). The models also predict that an increase in the initial inequality will cause a higher degree of income inequality with a higher degree of social disorder in all types of capitalist societies.

The available empirical data do not refute these predictions; hence, these models explain these particular traits of capitalist societies. Therefore, there is no reason to reject these models and we may accept the epsilon-omega-sigma theories of the capitalist system at this stage of our research. They are able to explain why the capitalist system operates with excess inequality and

social disorder, whereas in previous chapters they also explained the other disease of capitalism: why it operates with excess labor supply.

Up to now, the three partial theories epsilon, omega, and sigma have been able to explain the empirical regularities 1–4 that were established in chapter 2. In addition, as shown in this chapter, these theories have explained a new empirical regularity: the inequality and social disorder differences between the First World and the Third World, taken separately. Now it is time to turn to the construction of the unified theory of capitalism.

Appendix

Comparative Statics: Mathematical Proofs

The core of the static general equilibrium models of epsilon, omega, and sigma theories refers to the capitalist sector only. The core contains two equations and two endogenous variables: the nominal exchange rate (P_e) and the quantity of labor demanded (D_h). They were represented by equations (4.13), (5.6), and (6.9) in chapters 4, 5, and 6. These equations are identical. Comparative statics were derived graphically in the text; now the mathematical proof is presented in this Appendix. The equations are reproduced here just for convenience:

$$D_h = H\left(P_e; P_h, z^*; K_h\right), H_1 > 0, H_2 < 0, H_3 > 0, H_4 > 0$$

$$S_m = \mathrm{M}\left(P_e, D_h; P_h, P_h^*\right), \textit{where } M_i > 0, \text{ all i}$$

$$\mathrm{D}_h < \mathrm{D}_h^* < \mathrm{S}_\mathrm{h}.$$

In order to derive beta prepositions, we can follow the standard mathematical procedure. Partial differentiation of both equations with respect to each exogenous variable are applied and the system of equations so constructed is solved by using Cramer's rule.

Effect of changes in z:*

$$\frac{dD_h^\circ}{dz^*} = H_1 \frac{dP_e^\circ}{dz^*} + H_3$$

$$\frac{dS_m^\circ}{dz^*} = M_1 \frac{dP_e^\circ}{dz^*} + M_2 \frac{dD_h^\circ}{dz^*}$$

$$H_1 \frac{dP_e^\circ}{dz^*} - \frac{dD_h}{dz^*} = -H_3$$

$$M_1 \frac{dP_e^\circ}{dz^*} + M_2 \frac{dD_h^\circ}{dz^*} = 0.$$

$$\frac{dP_e^{\circ}}{dz^*} = \frac{\begin{vmatrix} -H_3 & -1 \\ 0 & M_2 \end{vmatrix}}{\begin{vmatrix} H_1 & -1 \\ M_1 & M_2 \end{vmatrix}} = \frac{-H_3 M_2}{H_1 M_2 + M_1} = \frac{-(+)(+)}{(+)(+)+(+)} = (-)$$

$$\frac{dD_h^{\circ}}{dz^*} = \frac{\begin{vmatrix} H_1 & -H_3 \\ M_1 & 0 \end{vmatrix}}{\begin{vmatrix} H_1 & -1 \\ M_1 & M_2 \end{vmatrix}} = \frac{H_3 M_1}{H_1 M_2 + M_1} = \frac{(+)(+)}{(+)(+)+(+)} = (+)$$

Effect of changes in S_m:

$$\frac{dD_h^{\circ}}{dS_m} = H_1 \frac{dP_e^{\circ}}{dS_m}$$

$$\frac{dS_m^{\circ}}{dS_m} = M_1 \frac{dP_e^{\circ}}{dS_m} + M_2 \frac{dD_h^{\circ}}{dS_m}$$

$$H_1 \frac{dP_e^{\circ}}{dS_m} - \frac{dD_h^{\circ}}{dS_m} = 0$$

$$M_1 \frac{dP_e^{\circ}}{dS_m} + M_2 \frac{dD_h^{\circ}}{dS_m} = 1.$$

$$\frac{dP_e^{\circ}}{dS_m} = \frac{\begin{vmatrix} 0 & -1 \\ 1 & M_2 \end{vmatrix}}{\begin{vmatrix} H_1 & -1 \\ M_1 & M_2 \end{vmatrix}} = \frac{1}{H_1 M_2 + M_1} = \frac{(+)}{(+)(+)+(+)} = (+)$$

$$\frac{dD_h^{\circ}}{dS_m} = \frac{\begin{vmatrix} H_1 & 0 \\ M_1 & 1 \end{vmatrix}}{\begin{vmatrix} H_1 & -1 \\ M_1 & M_2 \end{vmatrix}} = \frac{H_1}{H_1 M_2 + M_1} = \frac{(+)}{(+)(+)+(+)} = (+)$$

Effect of changes in K_b:

$$\frac{dD_b^\circ}{dK_b} = H_1 \frac{dP_e^\circ}{dK_b} + H_4 \frac{dK_b}{dK_b}$$

$$\frac{dS_m^\circ}{dK_b} = M_1 \frac{dP_e^\circ}{dK_b} + M_2 \frac{dD_b^\circ}{dK_b}$$

$$H_1 \frac{dP_e^\circ}{dK_b} - \frac{dD_b^\circ}{dK_b} = -H_4$$

$$M_1 \frac{dP_e^\circ}{dK_b} + M_2 \frac{dD_b^\circ}{dK_b} = 0.$$

$$\frac{dP_e^\circ}{dK_b} = \frac{\begin{vmatrix} -H_4 & -1 \\ 0 & M_2 \end{vmatrix}}{\begin{vmatrix} H_1 & -1 \\ M_1 & M_2 \end{vmatrix}} = \frac{-H_4 M_2}{H_1 M_2 + M_1} = \frac{-(+)(+)}{(+)(+)+(+)} = (-)$$

$$\frac{dD_b^\circ}{dK_b} = \frac{\begin{vmatrix} H_1 & -H_4 \\ M_1 & 0 \end{vmatrix}}{\begin{vmatrix} H_1 & -1 \\ M_1 & M_2 \end{vmatrix}} = \frac{H_4 M_1}{H_1 M_2 + M_1} = \frac{(+)(+)}{(+)(+)+(+)} = (+)$$

The effect of changes in each exogenous variable upon the rest of the endogenous variables is obtained just by implication from the solutions of the core. Because epsilon, omega, and sigma static models have different endogenous variables, the effect of each exogenous variable upon the endogenous variables will also be different.

Notes

3 Standard and Classical Economics

1. While the assumption of two social classes in the neoclassical and Keynesian theories above was introduced as an auxiliary one in order to generate models, this class structure is a primary assumption of the classical theory.
2. Roemer (1982) has developed a general theory of exploitation including the credit market. In his model, surplus labor must generate not only profits but also interest payments that go to banks. This is theoretically correct; empirically, however, profits are the most important component of surplus labor, that is, profits are much larger than interest incomes in national income accounts.

4 The Epsilon Society

1. A *limitational factor* is defined as follows: an increase in its quantity is a necessary but not sufficient condition for an increase in the level of output. This definition implies that this factor of production cannot be substituted by the others. In the technology assumed here, labor and capital can be substituted one for the other, but there is no substitution between input C and the combined factor labor-capital.
2. Dimensionally, the ratio c/z^* is a pure number. The condition that $(c/z^*) < 1$ ensures that the technological system is productive. If a firm exchanges one unit of good B for two units of good C as inputs, which help to produce less than one unit of good B, the only thing the firm has done is to run down its stock of good B. For the system to be productive these two units of good C should help to produce, say, four units of good B. Then the firm starts with one unit of good B and ends up with four units of it.
3. Under this system of flexible exchange rates, a particular rule for the functioning of the foreign exchange market can be assumed. Firms export commodity B and get foreign exchange, which they sell to the central bank and get domestic currency; then firms use domestic money to buy back foreign currency, which they use to import commodity C. Transaction costs in these

operations are zero. In this oversimplified model, the same firms are the actors behind demand and supply in the foreign exchange market. This market looks artificial because there is only one domestic good. In a multi-good economy, in which imports include production inputs and final goods, the agents that import will be different from those that export, and the equilibrium in the foreign exchange market will require to solve a coordination problem, through markets, because demand and supply conditions are determined independently. However, the conditions of equilibrium will be similar. Problems of foreign debt are ignored in this model.

5 The Omega Society

1. The sample includes Japan, Denmark, Germany, the United States, France, Sweden, Spain, and Portugal for the First World, and Mexico, Brazil, Panama, Costa Rica, Peru, Colombia, El Salvador, the Dominican Republic, Ecuador, and Bolivia for Latin America (ILO, World Employment Report 1997, Tables 2A, 2D, and 2E). (Calculations made by the author.)

Bibliography

Volumes 1 and 2

Acemoglu, D., Johnson S., & Robinson, J. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, 91 (5), 1369–1401.

Aeschbach-Hertig, W. (2007). Rebuttal of "On Global Forces of Nature Driving the Earth's Climate: Are Humans Involved?" by L. F. Khilyuk and G. V. Chilingar. *Environmental Geology*, 52, 1007–1009.

Alesina, A., & Perotti, R. (1996). Income Distribution, Political Instability, and Investment. *European Economic Review*, 40, 1203–1228.

Alesina, A., et al. (2003). Fractionalization. *Journal of Economic Growth*, 8, 155–194.

Atkinson, A. (1996). Seeking to Explain the Distribution of Income. In John Hills (Ed.). *New Inequalities. The Changing Distribution of Income and Wealth in the United Kingdom*. Cambridge, UK: Cambridge University Press.

Banfield, E. C. (1958). *The Moral Bases of a Backward Society*. New York: The Free Press.

Barro, R. (1997). *Macroeconomics* (5th Ed.). Cambridge: MIT Press.

Barro, R., & Lee, J. (2013). A New Data Set of Educational Attainment in the World. *Journal of Development Economics*, 104, 184–198.

Barro, R., & Sala-i-Martin, X. (2004). *Economic Growth* (2nd Ed.). Cambridge: The MIT Press.

Baumgärtner, S. (2004). The Inada Conditions for Material Resource Inputs Reconsidered. *Environmental and Resource Econo*mics, 29 (3), 307–322.

Becker, G. (1971). *The Economics of Discrimination* (2nd Ed.). Chicago, IL: Chicago University Press.

Blanchard, O. (2009). *Macroeconomics* (5th Ed.). Upper Saddle River, NJ: Prentice Hall.

Bonaiuti, M. (ed.). (2011). *From Bioeconomics to Degrowth: Georgescu-Roegen's New Economics in Eight Essays*. New York: Routledge.

Boulding, K. (1976). The Great Laws of Change. In Tang, A., Westfield, F., and Worley, J. (Ed.). *Evolution, Welfare, and Time in Economics*. Lexington, MA: Lexington Books.

Bourguignon, F. (2000). Crime, Violence and Inequitable Development. In *Annual World Bank Conference on Development Economics 1999*. Washington DC: The World Bank.

Bowles, S. (1985). The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian, and Marxian Models. *American Economic Review*, 75 (1), 16–36.

Carbaugh, R. 2011. *International Economics*. (13th Ed.). Mason, OH: Cengage Learning.

Chen, Y. et al. (2013). Evidence on the Impact of sustained Exposure to Air Pollution on Life Expectancy from China's Huai River Policy. *Proceedings of the National Academy of Sciences of the Unites States of America* (PNAS), 110 (32), August 6, 12936–12941.

Chilingar, G. V., Sorokhtin, O. G., & Khilyuk, L. F. (2008). Response to W. Aeschbach-Hertig Rebuttal of "On Global Forces Nature Driving the Earth's Climate. Are Humans Involved?" *Environmental Geology*, 54, 1567–1572.

Ciaian, P., & Kancs, A. (2011). Interdependencies in the Energy-Bioenergy-Food Price System: A Cointegration Analysis. *Resource and Energy Economics*, 33, 326–348.

Clugston, C. O. (2012). *Scarcity. Humanity's Final Chapter?* Bradenton, FL: Booklocker.com, Inc.

———. (2013). 21st Century NNR Scarcity: Blip or Paradigm Shift (An update of *Scarcity*). http://www.wakeupamerika.com/PDFs.

Currie, J. (2011). Inequality at Birth: Some Causes and Consequences. *American Economic Review*, 101 (3), 1–22.

Dadvand, et al. (2013). Maternal Exposure to Particulate Air Pollution and Term Birth Weight: A Multicountry Evaluation of Effect and Heterogeneity. *Environmental Health Perspectives*, 120 (12), 367–373.

Dalziel, N. (2006). *Historical Atlas of the British Empire*. London: Penguin Books.

Darity, W., & Nembhard, J. (2000). Racial and Ethnic Economic Inequality: The International Record. *American Economic Review*, 90 (2), 308–311.

Dasgupta, P., & Weale, M. (1992). On Measuring the Quality of Life. *World Development*, 20 (1) 119–131.

Davies, J., Sandstrom, S., Shorrocks, A., & Wolf, E. (2010). The Level and Distribution of Global Household Wealth. *Economic Journal*, 121, 223–254.

Deaton, A. (2013). *The Great Escape*. Princeton, NJ: Princeton University Press.

Deininger, K., & Squire, L. (1996). A New Data Set Measuring Inequality. *The World Bank Economic Review*, 10 (3), 565–591.

Deininger, K., & Squire, L. (1998). New Ways of Looking at Old Issues: Inequality and Growth. *Journal of Development Economics*, 57, 259–287.

Diamond, J. (1999). *Guns, Germs, and Steel: The Fates of Human Societies*. New York: W.W. Norton & Co.

Dominici, F. et al. (2002). Air Pollution and Mortality: Estimating Regional and National Dose-Response Relationships. *Journal of the American Statistical Association*, 97 (457), March, 100–111.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper Row.
Doyal, L., & Gough, I. (1991). *A Theory of Human Need*. London: Macmillan.
Drèze, J., & Sen, A. (2013). *An Uncertain Glory. India and Its Contradictions*. Princeton, NJ: Princeton University Press.
Emmott, S. (2013). *10 Billion*. London: Penguin Books.
Engerman, S., & Sokoloff, K. (1997). Factor Endowments, Institutions, and Differential Paths of Growth among New World Economies: A View of American Historians of the United States. In Stephen Haber (Ed.). *How Latin America Fell Behind: Essays on the Economic Histories of Brazil and Mexico, 1800–1914*. Stanford: Stanford University Press.
Etheridge, D., et al. (1998). Historical CO_2 Records from the Law Dome DE08, DE08–2, and DSS Ice Cores. In *Trends: A Compendium of Data on Global Change*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, TN. http://cdiac.ornl.gov/trends/co2/lawdome.html.
Fajnzylber, P., Lederman, D., & Loayza, N. (2002). What Causes Violent Crime. *European Economic Review*, 46, 1323–1358.
FAO (Food and Agricultural Organization, UN). (2009). *How to Feed the World in 2050*. www.fao.org.
Figueroa, A. (2008). Competition and Circulation of Economic Elites. *The Quarterly Review of Economics and Finance*, 48, 263–273.
———. (2009) *Capitalist Development and the Peasant Economy in Peru*. (Paperback Ed.). Cambridge, UK: Cambridge University Press.
———. (2010). Is Education Income Equalizing? Evidence from Peru. *CEPAL Review*, 102, April, 113–133.
———. (2011). Income Inequality and Credit Markets. *CEPAL Review*, 105, December, 37–51.
———. (2012). *The Alpha-Beta Method: Scientific Rules for Economics and the Social Sciences*. Buenos Aires, Argentina: Cengage Learning.
———. (2014). International Competition and Inequality: A Generalized Ricardian Model. Documento de Trabajo 378, Departamento de Economía, Pontificia Universidad Católica del Perú. Setiembre.
Fleurbaey, M. (2009). Beyond GDP: The Quest for a Measure of Social Welfare, *Journal of Economic Literature*, 47 (4), 1029–1075.
Forinash, K. (2010). *Foundations of Environmental Physics*. Washington, DC: Island Press.
Fukuyama, F. (2011). Left Out. *The American Interest*, 6 (3), January-February Issue.
Galor, O. (2011). *Unified Growth Theory*. Princeton, NJ: Princeton University Press.
Gardner, H. (1999). *Intelligence Reframed: Multiple Intelligences for the 21st Century*. New York: Basic Books.
Garraty, J. (1978). *Unemployment in History*. New York: Harper Colophon Books.
Gellner, E. (1983). *Nations and Nationalism*. Oxford, UK: Blackwell.

Georgescu-Roegen, N. (1960). Economic Theory and Agrarian Economics. *Oxford Economics Papers*, 12, 1–40.

——— (1971). *The Entropy Law and the Economic Process*. Cambridge, MA: Harvard University Press.

GFN (Global Footprint Network). (2013). Footprint Science. Data and Results. www.footprintnetwork.org. Accessed in September 2013.

Gollin, D. (2002). Getting Income Shares Right. *Journal of Political Economy*, 110 (2), 458–474.

Graubart, K. (2004). Hybrid Thinking: Bringing Postcolonial Theory to Colonial Latin American History. In E. Zein-Elabdin and S. Charusheela (Ed.). *Postcolonialism Meets Economics*. London: Routledge.

Hall, G., & Patrinos, A. (eds). (2005). *Indigenous Peoples, Poverty, and Human Development in Latin America*. New York: Palgrave Macmillan.

——— (eds). (2012). *Indigenous Peoples, Poverty, and Development*. Cambridge, UK: Cambridge University Press.

Hawking, S. (1996). *A Brief History of Time* (Updated and expanded Ed.). New York: Bantam Books.

Hertz-Picciotto, I., et al. (2007). Early Childhood Lower Respiratory Illness and Air Pollution. *Environment Health Perspectives*, 115 (10), 1510–1518.

Hirschman, A. (1973). The Changing Tolerance for Income Inequality in the Course of Economic Development. *Quarterly Journal of Economics*, 87 (4), 544–566.

Hudson, R. A. (1996). *Sociolinguistics*. Cambridge, UK: Cambridge University Press.

IHME (Institute for Health Metrics and Evaluation). (2012). Project Global Burden of Disease 2010: http://www.healthmetricsandevaluation.org/gbd/research/project/global-burden-diseases-injuries-and-risk-factors-study-2010.
The Report is published in a special issue of the British medical journal *The Lancet,* 380, December 15, 2012, pp. 2163–2260. Available online at: http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(12)61729-2/fulltext.

ILO (International Labor Organization). (1997). *World Employment Report 1997*. Geneva, Switzerland.

———. (2000). *Panorama Laboral*. Lima, Peru.

———. (2002). *Women and Men in the Informal Economy: A Statistical Picture.* Geneva, Switzerland.

IPCC (Intergovernmental Panel on Climate Change). (2007). The Physical Science Basis. Geneva, Switzerland.

———. (2013). Climate Change 2013: The Physical Science Basis. http://www.ipcc.ch.

Jones, C. (1998). *Introduction to Economic Growth*. New York: W.W. Norton.

Jones, C., & Vollrath, D. (2013). *Introduction to Economic Growth.* (3rd Ed.). New York: W. W. Norton.

Kalecki, M. (1971) *Selected Essays on the Dynamics of the Capitalist Economy 1933–1970*. Cambridge, UK: Cambridge University Press.

Keynes, J. M. (1936) *The General Theory of Employment, Interest and Money*. London: Macmillan Cambridge University Press.

Kristoufek, L., Janda, K., & Zilberman, D. (2012). Correlations between Biofuels and Related Commodities before and during the Food Crisis: A Taxonomic Perspective. *Energy Economics*, 34, 1380–1391.

Krugman, P., & Wells, R. (2006). *Macroeconomics*. New York: Worth Publishers.

Kuhn, T. (1970). *The Structure of Scientific Revolutions* (2nd Ed.). Chicago: University of Chicago Press.

Lafforgue, G. (2008). Stochastic Technical Change, Non-Renewable Resource and Optimal Sustainable Growth. *Resource and Energy Economics,* 30 (4), 540–554.

Lewis, A. (1954). Economic Development with Unlimited Supplies of Labor. *The Manchester School of Economic and Social Studies*, 28, 139–191.

Li, H., Squire L., & Zou, H. F. (1998). Explaining International and Inter-Temporal Variations in Income Inequality. *Economic Journal,* 108 (1), 26–43.

Lucas, R. (1978). Unemployment Policy. *American Economic Review*, 68 (2), 353–357.

——— (1990). Why Doesn't Capital Flow from Rich to Poor Countries. *American Economic Review*, 80 (2), 92–106.

Maddison, A. (1995). *Monitoring the World Economy, 1820–1992*. Paris: Development Centre, OECD.

———. (2003). *The World Economy: Historical Statistics*. Paris: OECD. Full data appear in http://www.ggdc.net/maddison/historical_statistics/horizontal-file_03-2007.xls.

Marx, K. (1867). *Capital*. London: Allen and Unwin, 1938.

Markusen, J. (2002). *Multinational Firms and the Theory of International Trade*. Cambridge: MIT Press.

Mc Connell, R., et al. (2010). Childhood Incident Asthma and Traffic-Related Air Pollution at Home and School. *Environment Health Perspectives*, 118 (7), 1021–1026.

McEvedy, C. (1972). *The Penguin Atlas of Modern History (to 1815)*. London: Penguin Books.

Milanovic, B. (2010). All the Ginis Dataset. Web Page at World Bank: http://go.worldbank.org/9VCQW66LA0. Last date accessed October 18, 2010.

Mohai, P., Kweon, B. S., Lee, S., & Ard, K. (2011). Air Pollution around Schools Is Linked to Poorer Student Health and Academic Performance. *Health Affairs*, 30 (5), 852–862.

Mukhopadhyaya, K., & Forssell, O. (2005). An Empirical Investigation of Air Pollution from Fossil Fuel Combustion and Its Impact on Health in India during 1973–1974 to 1996–1997. *Ecological Economics,* 55 (2), 235–250.

Muller, E. (1997). Economic Determinants of Democracy. In M. Midlarsky (Ed.). *Inequality, Democracy, and Economic Development*. Cambridge, UK: Cambridge University Press.

Muller, R. (2008). *Physics for Future Presidents: The Science behind the Headlines*. New York: Norton.

Munich R. E. (2013). Topics Geo. Natural Catastrophes 2012. Analyses, Assessments, Positions. http://www.munichreamerica.com/pdf/Topics_GEO_2012_us.pdf.

Musgrove, P. (2007). The Dethronement of Income as a Cause of Health: An Essay. *Brazilian Journal of Mother and Child Health,* 7 (4), 461–466.

Neiman, B., & Karabarbounis, L. (2013). The Global Decline of the Labor Share. National Bureau of Economic Research, Working Paper 19136, June.
Nolan, P. (2006). Marx, Karl (1818–1883). In Clark, D. A. (Ed.). *The Elgar Companion of Development Studies*. Cheltenham, UK: Edward Elgar.
North, D. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
OECD (Organization for Economic Cooperation and Development) (2010). *PISA 2009. Results: Executive Summary*. www.pisa.oecd.org.
Okun, A. (1975). *Equality and Efficiency: The Big Trade Off*. Washington, DC: The Brookings Institution.
Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge: Harvard University Press.
Olson, M., & Landsberg, H. (eds) (1975). *The No-Growth Society*. London and New York: Frank Cass.
Orchard, L., & Stretton, H. (1997). Public Choice: Critical Survey. *Cambridge Journal of Economics*, 21, 409–430.
Parker, J., Akinbami, L., & Woodruff, T. (2009). Air Pollution and Childhood Respiratory Allergies. *Environment Health Perspectives,* 117 (1), 140–147.
Patsiurko, N., et al. (2013). Nation-State Size, Ethnic Diversity, and Economic Performance in the Advanced Capitalist Countries. *New Political Economy*, 18 (6), 827–844.
Persson, T., & Tabellini, G. (2000). *Political Economics. Explaining Economic Policy.* Cambridge: MIT Press.
Pimentel, D. (2006). Soil Erosion: A Food and Environmental Threat. *Journal of Environment, Development, and Sustainability*, 8 (1), 119–137.
Piketty, T. (2014). *Capital in the Twenty First Century*. Cambridge, MA: Harvard University Press.
Popper, K. (1968). *The Logic of Scientific Discovery*. London: Routledge.
Psacharopoulos, G., & Patrinos, H. (1994). *Indigenous People and Poverty in Latin America. An Empirical Analysis*. Aldershot: Avebury.
Rand, J., & Tarp, F. (2002). Business Cycles in Developing Countries: Are They Different? *World Development*, 30 (12), 2071–2088.
Ratey, J. (2002). *A User's Guide to the Brain*. New York: Vintage Books.
Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
Ricardo, D. (1821). *On the Principles of Political Economy and Taxation*. Cambridge, UK: Cambridge University Press, 1951.
Rodriguez, J. (2013). Diferencias en ingresos laborales entre puestos de trabajo asalariados y puestos de trabajo autogenerados en el Perú, 2007–2011. *Desarrollo y Sociedad,* 171, junio, 123–163.
Rodrik, D. (2007). *One Economics, Many Recipes. Globalization, Institutions, and Economic Growth*. Princeton, NJ: Princeton University Press.
Roemer, J. (1982). *A General Theory of Exploitation and Class*. Cambridge, MA: Harvard University Press.
Rogoff, K. (1990). Equilibrium Political Budget Cycles. *American Economic Review*, 80 (1), 21–36.

Rousseau, J. (1755). *A Discourse on Inequality*. (Translation from French by Maurice Cranston.) New York: Penguin Books, 1984.

Salomon, J., et al. (2012). Healthy Life Expectancy for 187 Countries, 1990–2010: A Systematic Analysis for the Global Burden Disease Study 2010. *The Lancet*, 380, 2144–2162. www.thelancet.com

Samuelson, P. (1947). *Foundations of Economic Analysis*. New York: Atheneum, 1965.

Schneider, F., & Enste, D. (2000). Shadow Economics: Size, Causes, and Consequences. *Journal of Economic Literature*, 38, 77–114.

Schultz, T. (1975). The Value of the Ability to Deal with Disequilibria. *Journal of the Economic Literature*, 13, 827–846.

Searle, J. (1995). *The Construction of Social Reality*. New York: The Free Press.

Sen, A. (2006). Human Development Index. In D. A. Clark (Ed.). *The Elgar Companion to Development Studies*. Cheltenham: Edward Elgar.

Shapiro, C., & Stiglitz, J. (1984). Equilibrium Unemployment as a Worker Discipline Device. *American Economic Review*, 74 (3), 433–444.

Sheffield, P., Roy, A., Wong, K., & Trasande, L. (2011). Fine Particulate Matter Pollution Linked to Respiratory Illness in Infants and Increased Hospital Costs. *Health Affairs*, 30 (5), 871–878.

Silva, N. (2001). Race, Poverty, and Social Exclusion in Brazil. In E. Gacitúa, C. Sojo, and S. Davis (Ed.). *Social Exclusion and Poverty Reduction in Latin America and the Caribbean*. Washington, DC: The World Bank.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. New York: Random House, 1937.

Solow, R. (1974). Intergenerational Equity and Exhaustible Resources. *Review of Economic Studies,* 4, 29–45.

Stavins, R. (2011). The Problem of the Commons: Still Unsettled after 100 Years. *American Economic Review, Papers in Honor of the Centenary of the AER,* 101 (1), 81–108.

Stewart, F. (2001). *Horizontal Inequalities: A Neglected Dimension of Development*. In WIDER, Annual Lecture 5 presented at United Nations University, Helsinki.

——— (ed.) (2008). *Horizontal Inequalities and Conflict Understanding: Group Violence in Multiethnic Societies*. New York: Palgrave Macmillan.

Stiglitz, J. (2002). Information and the Change in the Paradigm in Economics. *The American Economic Review*, 92 (3), 460–501.

Stiglitz, J., Sen, A., & Fitoussi, J. (2010). *Mismeasuring Our Lives: Why GDP Doesn't Add Up*. New York: The New Press.

Telles, E. (1993). Urban Labor Market Segmentation and Income in Brazil. *Economic Development and Cultural Change*, 41 (2), 231–249.

UNCTAD (United Nations Conference on Trade and Development). (2006). *Trade and Development Report*. Geneva: United Nations Publication.

———. (2009). *Trade and Development Report.* New York and Geneva: United Nations Publication.

UNDP (United Nations Development Program). (2004). *Human Development Report 2004*. New York.

Walras, L. (1883). *Elements of Pure Economics*. New York: A. Kelly Publishers, 1954.

Wesseling, H. L. (2004). *The European Colonial Empires* 1815–1919. London: Pearson.

Wilson, E. (1998). *Consilience: The Unity of Knowledge*. New York: Alfred Knopf.

Wolff, E. (1998). Recent Trends in the Size Distribution of Household Wealth. *Journal of Economic Perspectives,* 12 (3), 131–150.

World Bank. (2001). *World Development Report 2000/2001: Attacking Poverty*. Washington, DC.

World Bank Development Indicators. (2010). http://data.worldbank.org/data-catalog/world-development-indicators. Particularly:

Infant and Under-5 Mortality Rates of World Development Indicators, Web Page at World Bank: http://databank.worldbank.org/Data/Views/VariableSelection/SelectVariables.aspx?source=World%20Development%20Indicators# Last date of revision: Not indicated. Last date accessed: 06/05/2013.

Life Expectancy database of World Development Indicators, Web Page at World Bank: http://databank.worldbank.org/Data/Views/VariableSelection/SelectVariables.aspx?source=World%20Development%20Indicators#. Last date of revision: Not indicated. Last date accessed: 06/05/2013.

Yim, S., & Barret, S. (2012). Public Health Impacts of Combustion Emissions in the UK. *Environmental Science and Technology*, 46 (8), 4291–4296.

Index

Subjects

Countries (excluding mentions in tables)

Authors (Volume I only)