# Inflationary Expectations
and the
# Natural Rate Hypothesis

Mats Persson

EFI THE ECONOMIC RESEARCH INSTITUTE
STOCKHOLM SCHOOL OF ECONOMICS

A Dissertation for the Doctor's Degree
in Economics
Stockholm School of Economics 1979



UDK 330.101.541
519.86
ISBN 91-7258-112-3
Åbergs Offset, Stockholm 1979

# Inflationary Expectations
# and the
# Natural Rate Hypothesis

# EFI

The Economic Reseach Institute at the Stockholm School of Economics
Address: Sveavägen 65, Box 6501, S-113 83 Stockholm, tel. 08-736 01 20

---

**Basic Orientation**

The Economic Research Institute (EFI), at the Stockholm School of Economics is devoted to the scientific study of problems in management science and economics. It provides research facilities for scholars belonging to these diciplines, training advanced students in scientific research. The studies to be carried out by the Institute are chosen on the basis of their expected scientific value and their relevance to the research programs outlined by the different sections of the Institute. The work of the Institute is thus in no way influenced by any political or economic interest group.

**Research Divisions:**

A Management and Personnel Administration
B Finance and Accounting
C Managerial Economics
D Marketing Management – Strategy and Economic Structure
F Public Administration
G Economic Geography
I Applied Information Processing
P Economic Psychology – Social and Individual Behavior
S Macro Economics

**Independent Research Program:**

Program for Studies of Participation and Organizational Development
Division for Applied Research in Business Administration

---

Additional information about research in progress and published reports is described in our project catalogue. The catalogue can be ordered directly from The Economic Research Institute, Box 6501, S-113 83 Stockholm, Sweden.

# PREFACE

The research leading to this thesis was initiated in 1974 under the auspices of Erik Lundberg, who wanted me to study the role of inflationary expectations for the determination of interest rates. During the years, the emphasis of the work has formed a random walk between different fields of macroeconomics, and has finally stabilized in the form of the present book. Throughout this process my debt to friends and colleagues has accumulated. In particular, I am grateful for the privilege of having had Karl-Göran Mäler as my thesis adviser; his clear insight and generous help have been indispensable for my work. The book has further benefited from valuable and constructive comments by Anders Björklund, Sören Blomquist, Peter Englund, Håkan Lyckeborg, Johan Myhrman, Erik Ruist, Lars Svensson, and Staffan Viotti. Also, I wish to thank Margareta Blomberg, Kerstin Niklasson and Monica Peijne for typing and re-typing the successive versions of this book.

To all of these, and to Bankforskningsinstitutet, which has provided financial support: my sincere gratitude.

Stockholm, April 26, 1979

Mats Persson

# CONTENTS

# 1. INTRODUCTION

The ultimate purpose of macroeconomic theory is to analyze the causes and interactions of inflation and unemployment. As a response to the pressure of economic reality, macroeconomic theory has developed rapidly during the last decade, and fruitful results have been gained along several lines of research. One of these lines, which is mainly associated with Milton Friedman, Edmund S. Phelps, Robert E. Lucas and their followers, has emphasized the role of expectations; when studying phenomena like inflation and unemployment, expectations has been the central concept to which most of the results have been related. Another line of research has been associated with Robert Clower, Axel Leijonhufvud, Robert Barro and Herschel Grossman, and their followers. It has emphasized price rigidities and the phenomena occurring in rationed markets, and it has built a "Keynesian" theory on these micro-foundations.

In this thesis I have chosen to approach the basic macroeconomic problems along the former lines, i.e. with the emphasis on inflationary expectations rather than on price rigidity. This is not because I believe this is the only, or even the "best", way to analyse macroeconomic phenomena; both approaches are rich and fruitful enough to deserve full attention. Since they cannot yet be meaningfully integrated, you have to choose one of them - and the problems of expectations formation seemed challenging indeed.

The reasons for choosing the particular subject I have done is thus that inflationary expectations and the natural rate hypothesis are important questions that - together with disequilibrium phenomena - constitute the very heart of macroeconomics. These questions can be dealt with in a number of ways; I have chosen to study them within the framework of stochastic processes. The concept of expectations implies that there is uncertainty involved. With no uncertainty there is no place for expectations; then we know for sure. But in the real world uncertainty is often the driving force behind macroeconomic phenomena, which should therefore be studied appropriately, i.e. by stochastic methods. The points made in this thesis are therefore such that they refer to the *stochastic nature* of the models utilized; similar results are not to be obtained from ordinary deterministic models.

The application of stochastic methods to macroeconomic models has the consequence that the analysis becomes fairly complicated. Therefore we have to confine the analysis to the very simplest and most straightforward models. A limitation is thus that we deal only with models for a *closed economy*. The most realistic analysis of inflation in e.g. Sweden should perhaps employ the methods developed for the study of small open economies; for the present thesis, however, we have chosen to abstain from this permutation of the basic models.

There is another qualification that should be made: We study only long-run phenomena. The natural rate hypothesis states that monetary policy cannot affect real magnitudes, such as unemployment, *in the long run*. It is therefore reasonable to ask whether the natural rate hypothesis is of any relevance, for economic policy or otherwise. How interesting are conclusions that apply only to the long run, in which we are all dead?

There are, I think, two answers to that question. The first answer is mainly theoretical: Long-run phenomena are interesting in their own right, in the same sense as abstract concepts like existence and stability are interesting in general equilibrium theory. Long-run solutions provide some kind of consistency, or some kind of frame of reference, which is necessary if we later want to study more practical, short-run phenomena. The second answer regards practical economic policy. If the natural rate hypothesis turns out to hold, we know exactly that: We *cannot* reduce unemployment permanently by monetary policy without getting a hyperinflation. If we try to peg unemployment at a rate below the "natural" one, we are bound to get an accelerating inflation. From the practical and political point of view, a study of whether the natural rate hypothesis holds or not seems to be a piece of information as interesting as any.

The book consists of three main parts. *The first part*, comprising Chapters 2 and 3, is intended as a methodological introduction. Since the book is meant to be as self-contained as possible, and since the stochastic methods employed are perhaps not yet common knowledge among those dealing with macroeconomics, the two first sections of Chapter 2 consist of an introductory survey of the theory of stochastic processes. The basic concepts in time series analysis, such as stationarity and ARMA processes, are presented; these sections can be skipped by readers familiar with those concepts. The third section of Chapter 2 consists of a discussion of stationarity and its interpretation in terms of dynamic equilibrium. Chapter 3 contains a presentation of the economic models, and the methods for modelling of expectations, that are to be used. It gives the economic background for the time series approach employed in later chapters.

*The second part* of the book consists of Chapters 4 and 5, and deals with adaptive expectations, i.e. expectations where the expected inflation rate $\pi_t^e$ is expressed as a weighted sum of earlier inflation rates:

$$\pi_t^e = \sum_{j=1}^{n} w_j \pi_{t-j}.$$

This expectations mechanism has a long tradition in economic and econometric studies. In Chapter 4 is discussed the common practice of constraining the $w_j$ weights to sum to unity. It is shown that if we want to make a forecast $\pi_t^e$ that is optimal in the sense of minimizing the mean square error

$$V \equiv E[(\pi_t - \pi_t^e)^2],$$

and if $\pi_t$ follows a so-called stationary stochastic process, then the adaptive forecast should have $w_j$ weights that sum to less than unity. Thus, if people actually form adaptive expectations that are optimal, and if they believe that inflation follows a stationary process (which means, intuitively, that inflation cannot explode) then the econometric practice of imposing the constraint $\Sigma w_j = 1$ will form a misspecification that can create bias in the estimates.

The result that $\Sigma w_j$ should be less than unity has consequences for the estimation of econometric models involving expectations terms. It will also have theoretical implications, and in particular the question occurs whether the natural rate hypothesis holds for macro models with optimal adaptive expectations. In Chapter 5 is demonstrated how this can be analyzed within the framework of a very simple model, and it is shown that the natural rate hypothesis actually holds even if $\Sigma w_j < 1$.

*The third part* of the book, comprising Chapters 6 and 7, deals with the so-called rational expectations, a fairly new concept which has been used in macro models during the seventies only, and which has increasingly gained attention in favor of the old, adaptive expectations. Rational expectations means that the inflation rate (or the price level) expected by the economic agents to prevail at time t is set equal to the true, mathematical expectation, conditional on the previous history of the economic system:

$$\pi_t^e = E[\pi_t | \pi_{t-1}, \pi_{t-2}, \ldots].$$

It has often been claimed that the natural rate hypothesis and the assumption of rational expectations are equivalent; it is said that models with rational expectations always display natural rate properties. In Chapter 6 is shown that this is not true for a commonly used macro model, and in particular it is shown that for this model, the average value of real output is an increasing function of the variance in the authorities' money supply rule. This result is quite contrary to conventional macroeconomic wisdom, which advocates care and stability in the money supply.

In Chapter 7, finally, the practice of inserting expectations terms $\pi_t^e$ (or $p_t^e$ for the price level) in economic models is discussed. The insertion of such variables requires that so-called certainty equivalents exist, i.e. that the uncertain variables $\pi_t$ and $p_t$ can be substituted by known constants $\pi_t^e$ and $p_t^e$. With rational expectations it further requires that the certainty equivalents $\pi_t^e$ and $p_t^e$ are equal to the mathematical expectations of $\pi_t$ and $p_t$. In Chapter 7 is shown that rational expectations terms in economic models have no solid micro-foundations. Thus, "rational expectations" are not really rational after all.

# 2. TIME SERIES ANALYSIS — AN INTRODUCTORY SURVEY[1]

## 2.1 DEFINITIONS AND BASIC CONCEPTS

Since 1749, the astronomers have collected data on the level of solar activity, i.e. on the monthly number of sun-spots. These figures form the time series plotted in Figure 2.1:

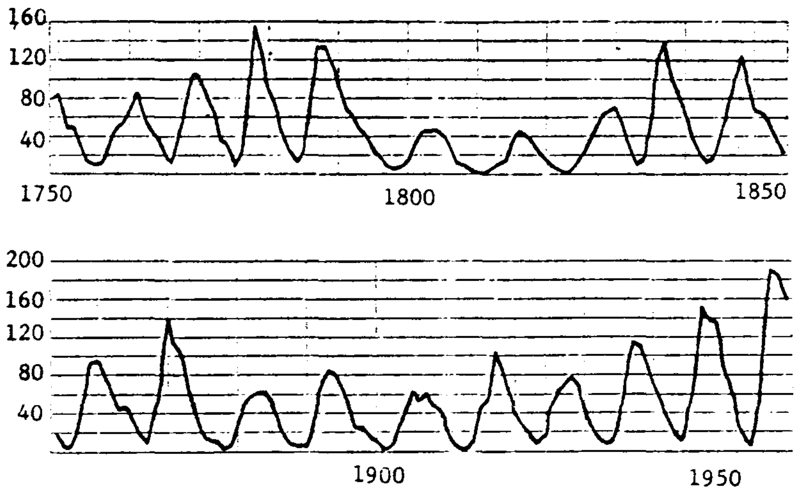


Figure 2.1: Monthly number of sunspots (Wölfer's numbers), 1749-1959

---

[1] In recent years a number of textbooks on time series analysis with economic applications have been published. This section draws heavily on Box and Jenkins (1970, 2nd ed. 1976), but very accessible introductions are also given by Nelson (1973) and Granger and Newbold (1977). An older textbook, which deals with more general and mathematical aspects of stochastic processes, with mostly engineering applications, is Yaglom (1962).

Similarly, the thickness of a 1 meter cotton thread is plotted in Figure 2.2, and although the horizontal axis does not indicate the passage of time, but the passage of a distance, these figures can also be said to form a "time series":

Figure 2.2: Thickness of a cotton thread, mm/10

As a third example we consider the change in the Swedish consumer price index since 1946 (Figure 2.3):

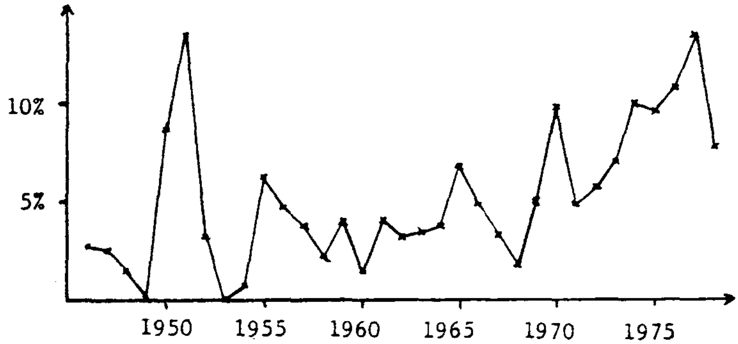


Figure 2.3: Yearly changes in the consumer price index, Sweden 1946-1978

Now, whether the world of ours is stochastic or deterministic is a highly philosophical question; with adequate knowledge of physical and social systems, we could perhaps figure out with certainty the number of sun-spots next month, or next year's inflation rate. With our limited knowledge, however, we can only make a more or less accurate guess, and the true outcome will in general differ from our guess. While lacking perfect knowledge, the analysis of several phenomena like the ones in Figures 2.1-2.3 above can be facilitated by regarding the time series as *stochastic processes*. This does not mean that such an approach is the "best" one for analyzing e.g. inflation; but it is a simple one, and for the present purpose it allows a fairly simple treatment of questions that would otherwise be of a prohibitive complexity.

A stochastic process is described by an ordered set of random variables, $X_t$. For the example of Figure 2.1, we let $X_t$ denote the number of sun-spots at time t (t = January 1749 - December 1959). Similarly, for Figure 2.2, we let $X_t$ denote the thickness of a thread at a point t in the interval [0, 1 meter]. In this book we will only consider discrete processes, i.e. we describe the process by the random variables

$$\{\ldots, X_{t-1}, X_t, X_{t+1}, \ldots\}.$$

This excludes the example with the thread above, where t obviously is a continuous variable. Such a limitation introduces no real restrictions on our analysis, since most economic data dealt with in this book are available only at discrete points in time, and since we could, if necessary, let the (discrete) time unit be arbitrarily small. Note that in the definition of a stochastic process there is nothing which says that the index t should be interpreted as calendar time, even if this is often the case (thereby the name time series). In the example with the cotton thread, t denoted distance; and numerous examples could be given where it denotes temperature, speed, income etc. Neither does $X_t$, nor t, have to be one-dimensional variables; in forestry, for example, methods of estimating land yield are utilized, where the stochastic variable $X_t$ denotes the volume (in cubic feet) of wood growing in the area of one square kilometer, while t is a two-dimensional variable denoting the coordinates of the different square kilometers. In the context of this book, however, $X_t$ will be a one-dimensional, real variable in discrete time, later interpreted as an economic variable like the rate of inflation, the price level, the rate of unemployment, or the level of real output.

The random variable is thus characterized by a distribution function, which gives the probability that $X_t$ be less than a given value x:

$$F_t(x) \equiv \text{Prob}(X_t \leq x).$$

Similarly, the realization of $X_t$ at two points of time, $t_1$ and $t_2$, is a two-dimensional random variable with

$$F_{t_1,t_2}(x_1,x_2) \equiv \text{Prob}(X_{t_1} \leq x_1,\ X_{t_2} \leq x_2).$$

For every number n, and for every set of dates $t_1$, $t_2$, ..., $t_n$ we have the n-dimensional distribution function

$$F_{t_1,..,t_n}(x_1,\ ..,\ x_n) \equiv \text{Prob}(X_{t_1} \leq x_1,\ ..,\ X_{t_n} \leq x_n).$$

Introducing the (unconditional) expectation E[.], we can write the mean or the expected value of the process at a time t as

$$\mu_t \equiv E[X_t] \equiv \int_{-\infty}^{\infty} X_t dF_t(x).$$

If the probability distribution has a density function $f_t(x)$, the mean can also be written

$$\mu_t \equiv E[X_t] \equiv \int_{-\infty}^{\infty} X_t f_t(x) dx.$$

Higher moments are also defined in the ordinary way; for example is the variance of $X_t$ given by

$$\mathrm{Var}[X_t] \equiv E[(X_t-\mu_t)^2].$$

Similarly, the covariance between the value of the variable at two points of time, say t and s, is defined by

$$\mathrm{Cov}[X_t,X_s] \equiv E[(X_t-\mu_t)(X_s-\mu_s)].$$

The process $\{\ldots, X_t, \ldots\}$, and the relations between $X_t$ and $X_s$, can take many forms, and it is necessary to introduce some simplifying assumption to make the analysis manageable. In fact, there are two such simplifications which are generally employed. They are adapted to different types of problems, and they are not mutually exclusive. The first one is the *Markov Chain Principle*, which can be stated as follows: For a general process, the value of $X_t$ depends on the value of $X_{t-1}$, $X_{t-2}$, $X_{t-3}$ etc. If we assume that $\{X_t\}$ is a Markov process, we assume that $X_t$ depends on $X_{t-1}$ only, that is, knowledge of the earlier history of the process (beyond t-1) does not add to our knowledge of the probability distribution of $X_t$. While of great importance in dealing with certain problems,[2] this is not the approach to be taken in this book. Instead, we will concentrate on the other simplifying assumption, that of *stationarity*.

*Definition 2.1:* If for all integers n and $\tau$, and for all sets $t_1, \ldots, t_n$, the n-dimensional random variable $X_{t_1+\tau}$, $X_{t_2+\tau}, \ldots, X_{t_n+\tau})$ has the same distribution as the n-dimensional random variable $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$, then the stochastic process $\{\ldots, X_t, \ldots\}$ is *strictly stationary*.

---

[2] See for example Cox and Miller (1965).

The definition says that if a process $\{X_t\}$ is strictly stationary, then the distribution function $F_t(x)$, as defined on the previous page, does not change with time. This means that we can dispense with the time subscript and write

$$F_t(x) = F_s(x) = F(x).$$

Thus a shift in the time scale does not change the distribution of the variable, which, among other things, implies that the mathematical expectation $E[X_t]$ is constant for all values of t. We then have, for all t and s,

$$\mu_t = \mu_s = \mu.$$

The assumption of stationarity also implies that the covariance between $X_t$ and $X_s$ depends only on the absolute value of the difference between t and s:

$$E[(X_t-\mu)(X_s-\mu)] \equiv \gamma_{|t-s|} \equiv \gamma_{t-s} \equiv \gamma_{s-t}.$$

We can thus write

$$\mathrm{Cov}[X_t, X_{t+\tau}] \equiv \gamma_\tau \qquad \tau = \ldots, -1, 0, 1, 2, \ldots$$

where $\gamma_\tau$ is independent of t. A process, the first two moments of which are independent of t, without necessarily satisfying the more restrictive Definition 2.1 above, is called weakly stationary, covariance-stationary, stationary of the second order, or just stationary. If the process is Gaussian, i.e. if, for all integers n, the variable $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ has an n-dimensional normal distribution, then weak stationarity obviously implies strict stationarity since, for normally distributed random variables, the whole probability distribution is completely characterized by the first two moments. For almost all practical purposes, or at least for the present ones, it is not necessary to deal with strict stationarity; weak stationarity is a sufficient concept, and we thus state:

*Definition 2.2:* If $E[X_t]$ is independent of t, and if $\text{Cov}[X_t, X_{t+\tau}]$ only depends on the absolute value of $\tau$, then the time series $\{X_t\}$ is *stationary*.

What will $\gamma_\tau$, regarded as a function of $\tau$, look like for a stationary process? If we return to the inflation rates of Figure 2.3, and if we assume that the time series of that example is stationary,[3] we can estimate the covariance function by

$$\hat{\gamma}_\tau = \frac{1}{N} \sum_{t=1}^{T-\tau} (x_t - \hat{\mu})(x_{t+\tau} - \hat{\mu}) \qquad \tau = \ldots, -1, 0, 1, \ldots$$

with

$$\hat{\mu} = \frac{1}{N} \sum_{t=1}^{T} x_t$$

where $x_t$ denotes the observed realization at time t, and where N is the number of observations.[4] When depicting the covariance function for model identification purposes, the covariances are often "normalized" by dividing through by $\gamma_0$. The resulting variable is called the *autocorrelation* and is shown in Figure 2.4:

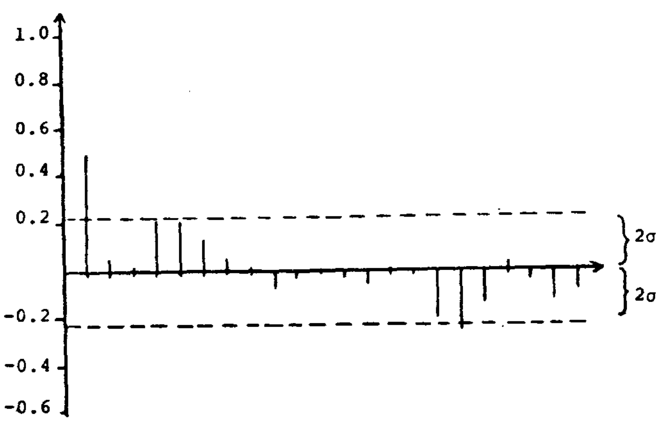


Figure 2.4: Autocorrelation function ($\rho_\tau \equiv \gamma_\tau/\gamma_0$) for Swedish inflation 1946-1978

---

[3] Whether or not such an assumption is appropriate for this particular case will be briefly discussed later in this chapter.

[4] Some authors have preferred to divide the sum of cross products by $N-\tau$ instead of $\tau$. This exposition follows the one of Box and Jenkins (1970); furthermore, in large samples the difference between $N-\tau$ and N will be of minor importance.

It is often difficult to obtain reliable results from time series with less than 40-50 observations; we can however form confidence intervals even for a series as short as the present one, to find that only the first autocorrelation $\rho_1$ is significantly different from zero.

We recall that for a stationary process $\gamma_\tau$ depends only on the absolute value of $\tau$, i.e. $\gamma_\tau = \gamma_{-\tau}$. Thus the function $\gamma_\tau$ is symmetric around $\tau = 0$, and $\gamma_\tau$ is completely determined for non-negative values of $\tau$. Therefore only the right part of the autocorrelation function is depicted in Figure 4. Furthermore, $\gamma_\tau$ reaches its maximum for $\tau = 0$, and the function damps out for large values of $\tau$:

$$\gamma_o > 0$$

$$\gamma_o > |\gamma_\tau|, \quad \tau \neq 0$$

$$\lim_{\tau\to\infty} \gamma_\tau = 0.$$

Another important property of the function is that the covariance matrix,

$$\Gamma \equiv \begin{pmatrix} \gamma_o & \gamma_1 & \gamma_2 & \cdots & \gamma_n \\ \gamma_1 & \gamma_o & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_2 & \gamma_1 & \gamma_o & \cdots & \gamma_{n-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\cdot & \cdot \\ \gamma_n & \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_o \end{pmatrix},$$

is positive definite. Similar properties also hold for the second moment not centered around the mean, i.e. for

$$r_\tau \equiv E[X_t \, X_{t+\tau}]. \qquad \tau = 0, 1, 2, \ldots$$

We thus have

$$r_o > 0$$

$$r_o > |r_\tau| \qquad \tau \neq 0,$$

and the matrix

$$R \equiv \begin{bmatrix} r_o & r_1 & r_2 & \cdots & r_n \\ r_1 & r_o & r_1 & \cdots & r_{n-1} \\ r_2 & r_1 & r_o & \cdots & r_{n-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_n & r_{n-1} & r_{n-2} & \cdots & r_o \end{bmatrix}$$

is symmetric and positive definite. Since $\gamma_\tau$ is defined by

$$\gamma_\tau \equiv E[(X_t-\mu)(X_{t+\tau}-\mu)],$$

while $r_\tau$ is defined by $r_\tau \equiv E[X_t\, X_{t+\tau}]$, we can easily compute the relation between $\gamma_\tau$ and $r_\tau$:

$$\gamma_\tau \equiv r_\tau - \mu^2. \qquad \tau = 0,\ 1,\ 2,\ \ldots$$

The covariance gives us information about how the values of $X_t$, for different points of time t, are related. If the process is Gaussian the covariance function, together with the expectation $E[X_t]$, gives us a complete characterization of the process $\{X_t\}$ in the sense that we, from the knowledge of $\gamma_\tau$ and $\mu$, can compute the distribution function for the process. Since the covariance function deals with how different values of $X_t$ are related over time, we say that it gives a representation of the process $\{X_t\}$ in the *time domain*. Another concept, which contains the same information as $\gamma_\tau$, is the *spectrum* of $\{X_t\}$, which is said to give a representation of the process in the *frequency domain*. The spectral approach is based on the idea that the time series is made up of sine and cosine waves with different frequencies, and the spectral mass function (or the spectral density function, if it exists) is a tool for illustrating which of these frequencies are dominating. Let $\{X_t\}$ be a stationary time series with a covariance function $\gamma_\tau$, $\tau = 0, 1, \ldots$ . Then there exists a unique mass distribution $F(\lambda)$ such that $\gamma_\tau$ and $F(\lambda)$ are connected by the one-to-one formula

$$\gamma_\tau = \int_{-\pi}^{\pi} e^{i\tau\lambda} dF(\lambda),$$

i.e. $\gamma_\tau$ is the so-called Fourier-Stieltjes transform of F. If $F(\lambda)$ has a density $f(\lambda)$, the formula can instead be written

$$\gamma_\tau = \int_{-\pi}^{\pi} e^{i\tau\lambda} f(\lambda) d\lambda.$$

The function $F(\lambda)$ is said to be the *spectral distribution function* of the process $\{X_t\}$.

Like there exist statistical methods for estimating the covariance function, there exist methods for estimating the spectrum of an observed time series. The estimated spectral density of the inflation rates in Figure 2.3 is shown in Figure 2.5:

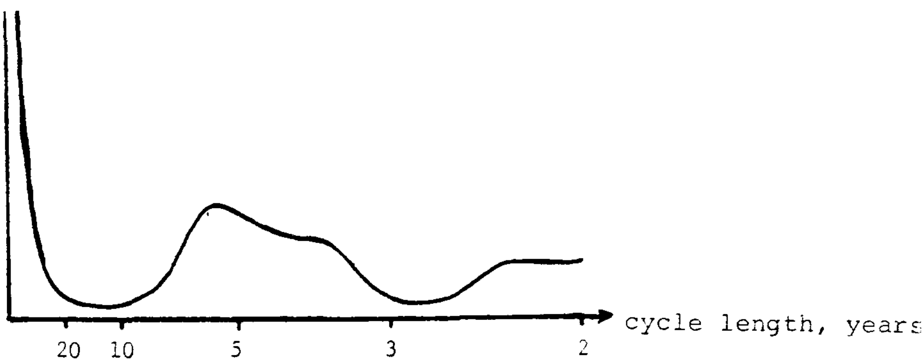


Figure 2.5: Spectral density for Swedish post-war inflation rates

Like in optics, the spectrum tells us which frequencies dominate in the oscillation. We see here, firstly, that there is a hump corresponding to the ordinary 5-year business cycle. Secondly, there is some spectral mass concentrated on the high frequencies (the shortest waves), and this is a statistical artifact due to the particular estimation program used; spectral mass corresponding to the very highest frequencies, that lie outside the diagram, is lumped together on this side of the spectrum. Thirdly, there is a lot of spectral mass concentrated on the lowest frequencies. This tells us that there might be some non-stationarity in the time series, or that our sampling period was too short for giving an accurate picture of the stochastic process. Returning to the original data in Figure 2.3 we see that the inflation rates show a tendency to increase over time. The estimation program interprets this as a long, low-frequency wave, which thus accounts for the high concentration of spectral mass in the left part of the diagram. For a longer time series, say 100 years or more, we would perhaps see that the tendency towards increasing inflation rates that we have observed during the last decades might in fact be the uphill part of a longer cycle which in due time turns down again. The estimated spectrum for such a time series would then perhaps show a distinct hump at cycle lengths of, say, 70 years, corresponding to such low-frequency "Kondratieff" cycles.[5] If we, on the other hand, had used quarterly data we would probably have been able to trace another hump in the diagram, corresponding to some once-a-year seasonality in the time series.

Although spectral analysis is an important tool in many applications, it has played a minor role in economic problems, which are instead generally studied in the time domain.[6] In

---

[5] A brief discussion of longer inflation series will be given at the end of this chapter.

[6] See however Granger and Hatanaka (1964) and Granger (1966). A standard textbook in spectral analysis is Jenkins and Watts (1968).

this book the main emphasis will consequently be on the covariance function, and the spectrum will be dealt with only occasionally.

## 2.2 AR AND MA PROCESSES

Eugen Slutsky presented in 1927 his famous explanation of the business cycle.[7] He pointed out that if we define a stochastic variable $X_t$ as a weighted sum of identically and independently distributed random variables, then the time series $\{X_t\}$ is most likely to display a cyclic behaviour. Let us write

$$X_t = \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + \dots + \beta_p\varepsilon_{t-p}$$

where $\{\varepsilon_t\}$ is a series of identically distributed random variables such that

$$E[\varepsilon_t] = 0$$

$$E[\varepsilon_t\varepsilon_{t-\tau}] = \begin{cases} \sigma_\varepsilon^2 & \text{for } \tau = 0 \\ 0 & \text{otherwise.} \end{cases}$$

This process $\{X_t\}$ is an example of the kind of time series Slutsky had in mind. Although the disturbances $\varepsilon_t$ are completely random, with no covariance and no cyclical behaviour, their weighted sum $X_t$ will in general display distinct "business cycles". The process $\{\varepsilon_t\}$ is often referred to as a *white noise* process, and $\{X_t\}$ is called *a moving-average process of order p*. Since $E[\varepsilon_t] = 0$, we have $E[X_t] = 0$ for the above example. A general form of a moving-average process of order p (an MA(p) process) allows for an arbitrary mean $E[X_t] = k$:

$$X_t = k + \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + \dots + \beta_p\varepsilon_{t-p}.$$

---

[7] Translated from Russian in Slutsky (1937).

There is nothing which restricts p to be finite; there are several examples of MA($\infty$) processes.

Now, let us regard another random variable $X_t$ which is formed by a weighted sum of earlier realizations $X_{t-1}$, $X_{t-2}$ etc. plus a white noise term:

$$X_t = m + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_q X_{t-q} + \varepsilon_t.$$

This is called an *autoregressive process of order q*, or an AR(q) process. Assuming that it is stationary, i.e. that $E[X_t] = E[X_{t-1}] = \dots = E[X_{t-q}]$, we can take the expectation of both members and obtain

$$E[X_t] = \frac{m}{1 - \Sigma\alpha_i}.$$

Every AR process can be written as an MA process. Consider for example the AR(1) process

$$X_t = m + \alpha X_{t-1} + \varepsilon_t. \tag{2.1}$$

Lagging one period yields

$$X_{t-1} = m + \alpha X_{t-2} + \varepsilon_{t-1}$$

which, substituted into (2.1), gives

$$X_t = m + \alpha m + \varepsilon_t + \alpha\varepsilon_{t-1} + \alpha^2 X_{t-2}. \tag{2.2}$$

Again lagging one period,

$$X_{t-2} = m + \alpha X_{t-3} + \varepsilon_{t-2}$$

can be substituted into (2.2) to obtain

$$X_t = m + \alpha m + \alpha^2 m + \varepsilon_t + \alpha\varepsilon_{t-1} + \alpha^2\varepsilon_{t-2} + \alpha^3 X_{t-3}.$$

After infinitely many substitutions we obtain the MA($\infty$) process

$$X_t = m(1+\alpha+\alpha^2+\ldots) + \varepsilon_t + \alpha\varepsilon_{t-1} + \alpha^2\varepsilon_{t-2} + \ldots \qquad (2.3)$$

Similarly, any MA process can be written on an AR form, which might be infinite. The choice of whether to write a particular process on AR or MA form is determined by which is the most convenient; for example, (2.1) is shorter than the equivalent (2.3); thus the former is to be preferred. In some cases, a process can be most conveniently written including both AR and MA terms, e.g.

$$X_t = k + \sum_{i=1}^{q} \alpha_i X_{t-i} + \sum_{j=1}^{p} \beta_j \varepsilon_{t-j} + \varepsilon_t,$$

whereby we talk of an ARMA(q,p) process.

Now, what do these AR and MA processes have to do with the concepts discussed in the previous section? There are three questions to be answered: *Firstly*, can an arbitrary time series $\{X_t\}$ be written on AR or MA form? *Secondly*, what are the stationarity conditions for an AR or MA process? *Thirdly*, can the covariance function be expressed in terms of the autoregressive or moving-average coefficients $\alpha_i$ and $\beta_j$?

The first of these questions can be approached by the so-called Decomposition Theorem, due to Wold (1954). He proved that any stationary process $\{X_t\}$ can be written as the sum of two mutually uncorrelated processes,

$$X_t = D_t + Y_t,$$

where $\{D_t\}$ is a linearly deterministic process[8] and $\{Y_t\}$ is an MA process. Now there are many examples of deterministic processes $\{D_t\}$, for example

$$D_t \equiv a \cdot \cos bt. \tag{2.4}$$

Assume that $\{D_t\}$ is given by (2.4); the process $X_t = D_t + Y_t$ would then be given by

$$X_t = a \cdot \cos bt + \sum_{j=1}^{p} \beta_j \varepsilon_{t-j} + \varepsilon_t. \tag{2.5}$$

Now we take the mathematical expectation of both members of (2.5) to obtain

$$E[X_t] = a \cdot \cos bt.$$

Since $E[X_t]$ obviously varies with t, $\{X_t\}$ cannot be stationary. For $\{X_t\}$ to be stationary, the deterministic process $\{D_t\}$ in the Wold Decomposition must be independent of time, i.e. a constant:

$$D_t = k.$$

Thus *every stationary stochastic process* can be written on the form

$$X_t = k + \sum_{j=1}^{p} \beta_j \varepsilon_{t-j} + \varepsilon_t$$

or on the equivalent AR form

$$X_t = m + \sum_{j=1}^{q} \alpha_j X_{t-j} + \varepsilon_t.$$

---

[8] A time series $\{D_t\}$ is said to be linearly deterministic if there exists a linear function of past and present values $d_t \equiv \delta + \delta_o D_t + \delta_1 D_{t-1} + \delta_2 D_{t-2} + \dots$ such that $E[(D_{t+1} - d_t)^2] = 0$.

Thus the fact that we confine our analysis to MA or AR processes does not imply any restriction, as long as we are satisfied with studying stationary processes. There are however some kinds of processes that we exclude from our analysis, the most important being perhaps the processes that have a deterministic wave movement underlying them, for example the ones with $\{D_t\}$ given by the cosine wave (2.4). In this latter case, the deterministic wave appears as a spike in the spectral mass distribution.

Practically all economic variables display time series without such spikes, i.e. having spectral densities which allow us to express them on AR or MA form, which is quite convenient for analysis and estimation purposes. Note that this does not rule out distinct cyclical movements in the time series; assuming that the spectral density exists does only mean that there is no *deterministic* sine or cosine wave in the process. As shown in the paper by Slutsky referred to at the beginning of this section, such a process might well display "business cycles", the only qualification being that the length of these cycles is not constant for all times but varies, however slightly, in a random manner.

Now, given an AR(q) or MA(p) process - how do we know whether or not it is stationary? For an MA process

$$X_t = k + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + \dots + \beta_p\varepsilon_{t-p} + \varepsilon_t \qquad (2.6)$$

we state, without proof, the following conditions for stationarity:

$$\begin{aligned} &|\Sigma\beta_j| < \infty \\ &|\Sigma\beta_j\beta_{j+\tau}| < \infty. \qquad \tau = 0, 1, 2, \dots \end{aligned} \qquad (2.7)$$

These conditions are trivially satisfied for MA(p) processes of finite order; only MA($\infty$) processes can be non-stationary. For example, this means that the process (2.3) is stationary if and only if $|\alpha| < 1$. In such a case,

$$\sum_{j=1}^{\infty} \alpha^j = \frac{1}{1-\alpha} - 1 < \infty.$$

If (2.7) are satisfied, we can take the expectation of both members of (2.6) and obtain

$$E[X_t] = k.$$

We can also compute $E[(X_t - E[X_t])(X_{t+\tau} - E[X_{t+\tau}])]$ which gives

$$\gamma_\tau = \sigma_\varepsilon^2 \sum_{j=o}^{p} \beta_j \beta_{j+\tau} \qquad \tau = 0, 1, \ldots, p \tag{2.8}$$

$$\gamma_\tau = 0 \quad \text{for} \quad \tau > p.$$

For an AR process the stationarity conditions are somewhat different. Given a time series represented by

$$X_t = m + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_q X_{t-q} + \varepsilon_t$$

we define the *characteristic equation* of $\{X_t\}$ as

$$1 - \alpha_1 z - \alpha_2 z^2 - \ldots - \alpha_q z^q = 0.$$

An AR(q) process is stationary if and only if all the roots $\zeta_i$ of the characteristic equation lie outside the unit circle in the complex plane, i.e. if and only if

$$|\zeta_i| > 1. \qquad i = 1, \ldots, q$$

We see, for example, that if the AR(1) process (2.1) is written on the equivalent MA($\infty$) form (2.3), the stationarity condition for MA processes state that $\alpha$ must be less than unity in absolute value. But the characteristic equation of (2.1) is

$$1 - \alpha z = 0,$$

the (single and real) root of which is greater than unity in absolute value if the absolute value of the coefficient $\alpha$ is < 1. Thus, the stationarity condition for the AR(1) process is equivalent to the one for the corresponding MA($\infty$) process.

As another example we could take the AR(2) process

$$X_t = -0.7X_{t-1} - 0.6X_{t-2} + \varepsilon_t. \tag{2.9}$$

The characteristic equation is then

$$1 + 0.7z + 0.6z^2 = 0$$

which has the two roots

$$\zeta_1 = -0.58 + 1.15i$$
$$\zeta_2 = -0.58 - 1.15i.$$

The position of these roots in the complex plane is shown in Figure 2.6:

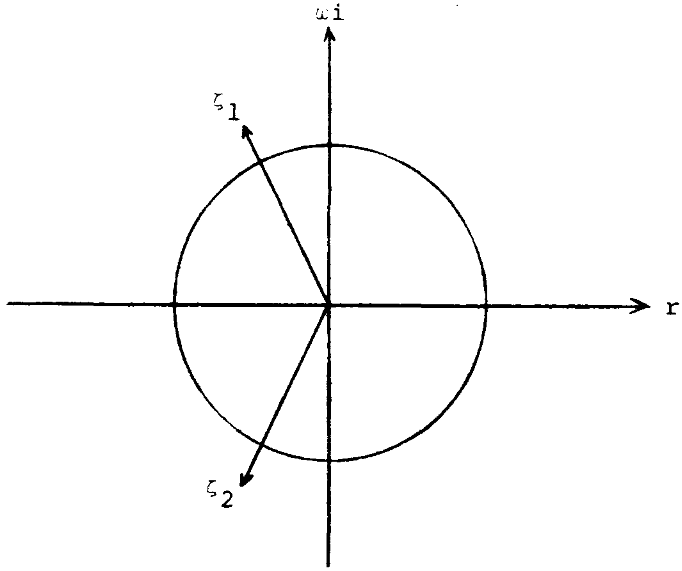


Figure 2.6: The roots of equation (2.9)

Obviously the process is stationary, since the roots lie well outside the unit circle.

Having established the stationarity of an AR(q) process, we can by some computation obtain the covariance function, expressed in terms of the autoregressive parameters:

$$\gamma_o = \alpha_1\gamma_1 + \alpha_2\gamma_2 + \dots + \alpha_q\gamma_q + \sigma_\varepsilon^2$$

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \cdot \\ \cdot \\ \cdot \\ \gamma_q \end{pmatrix} = \begin{pmatrix} \gamma_o & \gamma_1 & \cdots & \gamma_{q-1} \\ \gamma_1 & \gamma_o & \cdots & \gamma_{q-2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \gamma_{q-1} & \gamma_{q-2} & \cdots & \gamma_o \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_q \end{pmatrix}$$

$$\gamma_\tau = \alpha_1\gamma_{\tau-1} + \alpha_2\gamma_{\tau-2} + \dots + \alpha_q\gamma_{\tau-q} \qquad \tau > q.$$

These equations are sometimes called the *Yule-Walker equations*.

A stationary AR process is thus a process with all its characteristic roots outside the unit circle in the complex plane; the opposite case, that of a non-stationary, explosive process, has a characteristic equation with at least one root within the unit circle. An intermediate case which plays an important role in several economic contexts is that of the *random walk* - an AR(1) process[9] that has its only root $\zeta = 1$:

$$X_t = X_{t-1} + \varepsilon_t.$$

$E[X_t]$ is undefined for this process, as is $\gamma_\tau$. The random walk is thus non-stationary, and like an "ordinary", non-stationary process with roots strictly within the unit circle, it attains infinitely large values with probability one. But unlike such processes, which disappear far out to infinity, the random walk *returns to its point of departure* with probability one. For a visual impression of the different kinds of time series, the realizations of a stationary process, of a random walk and of an "ordinary non-stationary" process are shown in Figure 2.7:

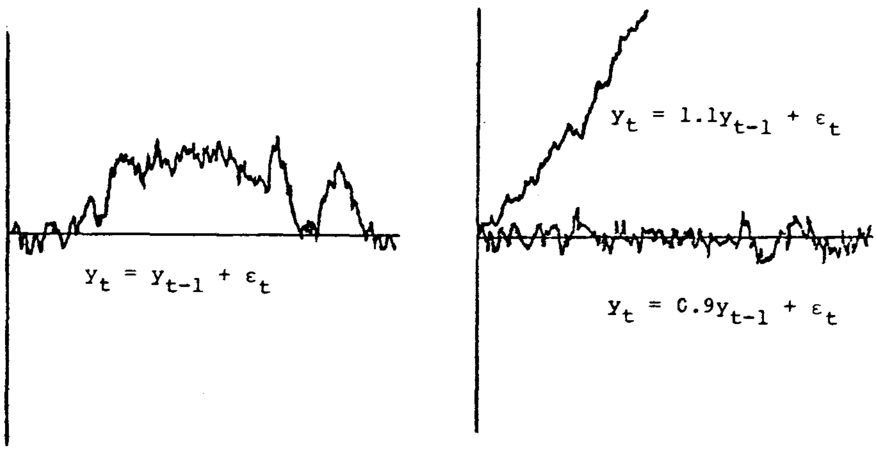


Figure 2.7: Simulations of AR processes with characteristic roots on, strictly outside, and strictly within the unit circle

[9] The equivalent MA(∞) form of the random walk is $X_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots$

Although this book mainly deals with theoretical aspects of time series analysis, and empirical questions play a very minor role, a few words should be mentioned about *estimation*. For a given time series, like the inflation rates in Figure 2.3, we can estimate the autocorrelation function as shown in Figure 2.4. Since there is a unique correspondence between the covariance and the AR or MA coefficients, the estimated $\hat{\rho}_\tau$ function can thus give a rough hint of what kind of AR or MA process that has generated the time series. This hint can be used as an initial estimate in the estimation procedure to obtain the $\alpha_i$ or $\beta_j$ coefficients. The particular estimation procedures employed are often different for the different computer programs available in the market; some programs make the estimates by an iterative, maximum-likelihood procedure, while others employ non-linear regression techniques minimizing the sum of squares of the residuals. When final estimates are obtained, there exists a number of tests of whether these coefficients are significantly different from zero, whether the resulting residuals are significantly uncorrelated etc.[10]

## 2.3 ON THE RATIONALE FOR STATIONARITY

Throughout this book, stationarity will be a basic concept. The analysis will deal mainly with stationary processes, and economic concepts like the natural rate hypothesis will be interpreted in terms of the stationarity properties of the time series formed by the economic variables. We will end this chapter by briefly discussing the justifications for confining the attention to the concept of stationarity.

---

[10] For a more detailed description of these procedures, see the textbooks referred to in footnote 1.

The *first* reason is that of *simplicity*. Stationary processes are easy to recognize, and the conditions for stationarity are easy to handle. This holds in particular if the time series can be represented as moving-average or autoregressive processes; we can then apply the tools demonstrated in the preceding section. The reason for studying stationary AR and MA processes in macro theory is thus the same as the reason for studying "stationary states" in (deterministic) growth theory; there is no alternative if one wants to obtain manageable results.

The quest for simplicity is thus a necessary, but not a sufficient reason for choosing a particular approach to study economic phenomena. The *second* reason for concentrating on stationarity is that it has a *meaningful economic interpretation*; it can be viewed in terms of dynamic equilibrium. There are of course many aspects of equilibrium, for example the equality of supply and demand. Another aspect refers to the concept of consistency; that the prices consumers and producers take into account when they make their decisions are the same as the ones appearing in the solution of the general equilibrium system, or that the inflationary process for which the agents form their expectations is the same as the one occurring as a result of these expectations. In dynamic models a third dimension of equilibrium occurs, namely that of constancy, steady-state, or stationarity. A model that repeats itself period after period is said to be in dynamic equilibrium. In deterministic models dynamic equilibrium means that all variables are constant;[11] in stochastic models the straightforward analogy of dynamic equilibrium is that all variables' *probability distributions* be constant.

---

[11] Sometimes however the constancy does not refer to the variables' *levels*, but to their *rates of change*. In a "stationary growth model" all magnitudes are increasing at a constant rate.

An obvious analogy is provided by deterministic macromodels. In static macromodels we can connect combinations of inflation and unemployment to form a Phillips curve. Making the model *dynamic*, we can analyze the *long-run* Phillips curve. This is done by investigating whether the (deterministic) paths of the inflation rate $\pi_t$ and the unemployment rate $u_t$ ultimately stabilize at some constant values $\pi$ and $u$. The long-run Phillips curve is then the curve connecting pairs of such constant values in the $(\pi,u)$-plane. In a *stochastic* model the variables never attain constant values, but we can take the average values $E[\pi_t]$ and $E[u_t]$ and compare these. The analogy to the deterministic case lies in the fact that we can compute these pairs of numbers, plot them in the $(E[\pi_t], E[u_t])$-plane and interpret the result as a long-run Phillips curve. For such an interpretation to be meaningful, $E[\pi_t]$ and $E[u_t]$ must be defined, and must be *constant over time*. And this means that we have to study whether the time series described by $\pi_t$ and $u_t$ are stationary.

There is a minor problem in this interpretation. It does not seem too unreasonable to extend the concept of dynamic equilibrium in deterministic models to the cases where variables are not really constant, but fluctuate in a stable manner, for example according to a cosine wave

$$Y_t = a \cdot \cos bt. \tag{2.10}$$

As long as the variables do not explode, such models could be said to be in a state of equilibrium. Assume now that we have a stochastic model of the same kind as the deterministic one, except for a disturbance term $\varepsilon_t$ which affects $Y_t$ additively:

$$Y_t = a \cdot \cos bt + \varepsilon_t. \tag{2.11}$$

If $Y_t$ as given by the deterministic model (2.10) is said to be in a state of dynamic equilibrium, then $Y_t$ as given by the stochastic model (2.11) could equally well be said to form an equilibrium process. In the time series terminology, however, the process (2.11) is non-stationary and will thus not be considered in our analysis. Now, cosine waves like the one above is only one example of recurrent, deterministic phenomena which cannot be analyzed in terms of stationary ARMA processes; another type of deterministic influence is "the political business cycle", stemming from the fact that the election period is deterministic.[12] Without denying the potential importance of such phenomena, I have chosen to disregard them for the present purpose.

For time series that can be represented by AR or MA processes, non-stationarity can take two forms: either the time series virtually explodes (like in the case with the process $X_t = 1.1\ X_{t-1} + \varepsilon_t$ in Figure 2.7 above), or it drifts away like the random walk, but returns with probability one. In the first case it is evident that the non-stationarity can be looked upon as a phenomenon of dynamic disequilibrium, while the second case is a little more doubtful. It is hard, however, to identify the random walk with a state of dynamic equilibrium, since the variable does not move in any stable manner around some well-defined average value.

Finally, it would be interesting to see how well the stationarity property conforms to observed reality, for example in the inflation time series. We recall that the Swedish inflation rates as depicted in Figure 2.3 above seemed fairly stationary, although there was perhaps a slight tendency for increases over time. Although the time series is quite short (the Box-Jenkins estimation procedures require at least 40-50 observations to yield any reliable estimates) we can estimate the ARMA process underlying it, carry out the tests for stationarity etc., to obtain the stationary MA(1) process

---

[12] Cf. Nordhaus (1975) and Lindbeck (1976).

$$\pi_t = 5.17 + 0.56\ \varepsilon_{t-1} + \varepsilon_t \qquad \sigma_\varepsilon^2 = 8.87 \tag{2.12}$$

To obtain better estimates, however, we need more observations. On the other hand, very ancient inflation figures are perhaps of minor relevance for today's economic policy. In particular, people's memory is limited, and nobody takes the time series from, say, the 19th century into account when forming expectations about tomorrow's inflation rate.

Just for the sake of curiosity, we display the Swedish yearly inflation rates between 1732 and 1977 in Figure 2.8:

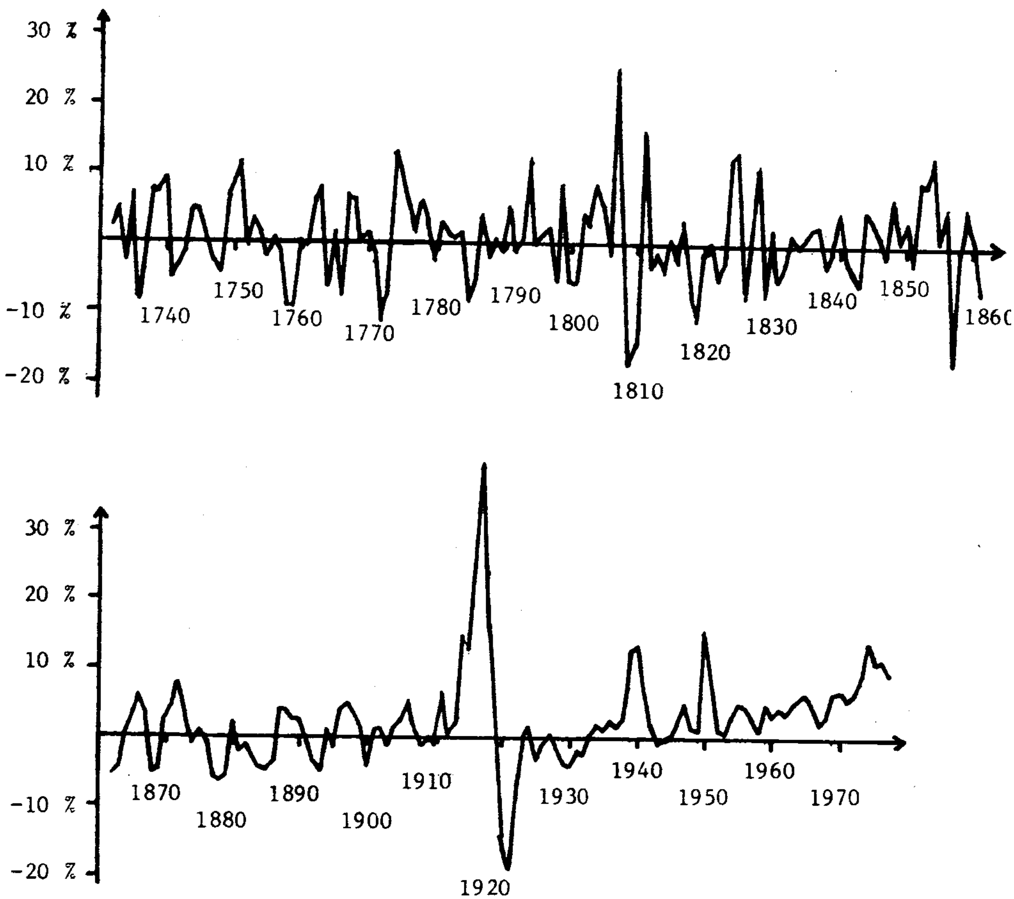


Figure 2:8: Yearly changes in the implicit GNP deflator, Sweden 1732-1977. The inflation rate at time t is defined by $\pi_t \equiv (P_{t+1}-P_t)/P_t$, where $P_t$ is the price leve l at time t.
Sources: Åmark (1921), Johansson (1959) and Swedish National Accounts.

This time series looks fairly stationary, and the rise in inflation rates during the 1970's is in this long perspective of a minor significance. Trying to estimate the spectrum of this process, we obtain the picture displayed in Figure 2.9:

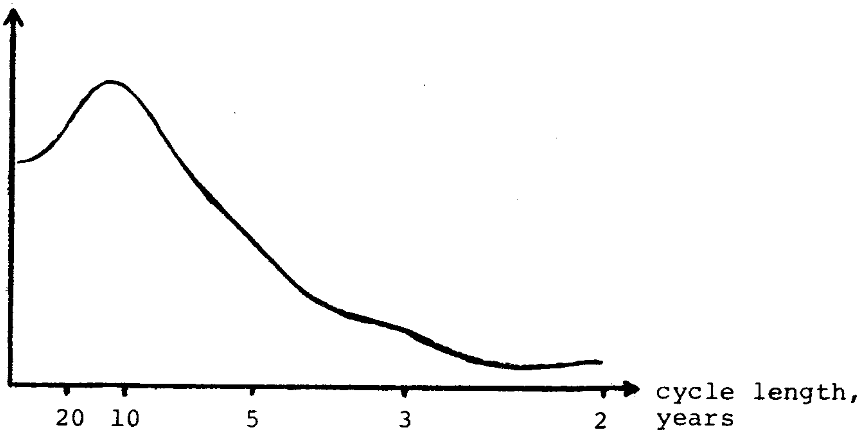


Figure 2.9: Estimated spectrum for Swedish inflation rates 1732-1977

Firstly we notice that the high concentration of spectral mass at the lowest frequences of Figure 2.5 has diminished because of the higher number of observations. Secondly there does not seem to be any long, Kondratieff waves in the series, and just a slight hint of a 3- or 4-year business cycle. Instead the dominating frequence seems to be an eight-year cycle, unheard of in the literature on business cycles. The explanation of this is simple, however: The great disturbances created by the First World War, and to a lesser extent by the 1808-1809 war, have hidden all other periodicities. Especially the eight-year cycle that occurred around the First World War dominates so overwhelmingly that hardly any other cycles are visible. To give a fair picture of the business cycle during the last two centuries, we should take these effects into account. However, this clearly is beyond the scope of this book; we simply display the crude estimate of Figure 2.9 as an illustration of a statistical method. In the same spirit we estimate the ARMA process underlying the Swedish inflation since 1732; it turns out to be the MA(1) process

$$\pi_t = 1.69 + 0.63\ \varepsilon_{t-1} + \varepsilon_t. \qquad \sigma_\varepsilon^2 = 38.7$$

Now this very long time series encompasses the old, rural Sweden of the 18th and the first half of the 19th century, as well as the industrial society emerging after, say, 1860. It is hardly surprising that estimating the ARMA processes for these separate subperiods yields quite different results. Thus we have for the period 1732-1860

$$\pi_t = 0.93 + \varepsilon_t \qquad \sigma_\varepsilon^2 = 42.90$$

with hardly any business cycles. For the period 1861-1977 we have instead

$$\pi_t = 0.93 + 0.92\ \pi_{t-1} - 0.29\ \pi_{t-2} - 0.26\ \varepsilon_{t-4} + \varepsilon_t$$

$$\sigma_\varepsilon^2 = 20.66$$

with a pronounced four-year cycle. These estimates should not be taken too seriously, however, since they are still blurred by the impact of the wars 1808-1809 and 1914-1918.

The fact that the inflationary process seems to have changed since the 18th century might of course be a sign of non-stationarity. On the other hand it could be argued that it is the same stationary process, but that it contains such long waves that our 246 observations are too few for detecting them. Such a discussion is hardly meaningful, however, and we should not go deeper into it, but just notice that inflation has not displayed any tendency to explode during the last 246 years, thereby lending some support to the hypothesis of stationarity.

# 3. EXPECTATIONS IN MACROECONOMICS

## 3.1 TWO SIMPLE MACRO MODELS

In this chapter we will investigate how inflation and unemployment can be studied within a time series framework. There exists a multitude of models connecting real and nominal variables in economics, and we will confine the analysis to two of them: the *Phillips Curve* and the *Aggregate Supply Function*. The reasons for dealing with these two models, which are not mutually exclusive, are that they are *simple*, thereby allowing us to concentrate on our main points concerning the stochastic properties of the models and the corresponding formation of expectations. Furthermore, they are *relevant*; these two models have been in the focus of almost all contemporary macroeconomic model-building, and macroeconomic debate, and therefore it seems appropriate to concentrate upon them.

### 3.1.1 The Phillips Curve

Irving Fisher (1926) was the first to investigate the relation between inflation and unemployment, a relation which was rediscovered by Phillips (1958). Three questions were quite naturally asked as soon as the Phillips curve became known:[1]

i) Which are the micro-foundations?

ii) Is the relationship stable?

iii) What implications does the relationship have for economic policy?

*The first question* was dealt with by Lipsey (1960) in terms of the Walras-Samuelson tâtonnement procedure, which assumes that the rate of increase in the relative price of a commodity is a function of the excess demand for that commodity. Regarding (nominal) wage inflation as a proxy for the increase in the relative price of labor, and the rate of unemployment as a proxy for (negative) excess demand for labor, such an interpretation would imply a Phillips-like relation. However, many problems arise: The increase in nominal wages is quite a bad proxy for the increase in the relative price of labor, especially since a Phillips relation can also be estimated if the change in nominal wages is substituted by the change in consumer price index. Further, the Walrasian auctioneer, swiftly adjusting prices before any trading takes place, is indeed a strange creature; the policy-minded economist would perhaps prefer some more realistic explanation, based on the behavior of firms and workers, of why price changes and unemployment rates are related in the way they are. Finally, the tâtonnement procedure cannot explain the distinct counter-clock-wise "loops" around the Phillips curve that had been observed.

By the term "micro-foundations" we usually mean the fact that an observed relationship like the Phillips curve can be derived from a basic model of maximizing behavior. Although Lipsey's attempt to explain the inflation-unemployment rela-

---

[1] Since there exist by now several good surveys of the history of the Phillips curve, we shall give only a brief account here. For a more detailed treatment, see e.g. R.J. Gordon (1976), Milton Friedman's Nobel Lecture (1977), and Santomero and Seater (1978).

tionship in terms of the Walras-Samuelson tâtonnement dealt with fundamental concepts in microeconomic theory, it did not provide micro-foundations in the stricter sense of stating explicitly which maximization model leads to the observed behavioral relation.

A more influential (and perhaps more successful) attempt to provide micro-foundations for the inflation-unemployment trade-off was launched by the publication of the so-called Phelps volume (1970). In that volume, which during the nine years since its first appearance has become a classic in its field, Holt (1970) derived a Phillips-like relationship out of workers' search for new jobs and acceptance of wage offers. Similarly, Phelps (1970) derived the relationship from a model of firms' optimal strategy for wage offerings in order to attract workers. In a third paper, Mortensen (1970) integrated both the behavior of workers, searching employment, and firms, trying to attract labor, thereby also obtaining a Phillips-like relation between inflation and unemployment.[2]

Two characteristic features of the papers in the Phelps volume should be mentioned. *Firstly*, they do not regard unemployment as a disequilibrium phenomenon, like in Lipsey's tâtonnement reasoning referred to above, or like in Keynesian or neo-Keynesian models (such as, for example Clower's (1965) and Barro-Grossman's (1976) models). Instead, they regard unemployment as a voluntary activity of rational agents optimally searching for better job opportunities. *Secondly*, most of "the new microeconomics" models incorporate *the natural rate hypothesis*, which brings us into the second of the questions on page 35 above, namely whether the Phillips relation is stable. The answer has to do with whether, and how, inflationary expectations should be modelled into the Phillips curve, which will be the main theme of this book.

---

[2] More recent, and more general treatments of the questions dealt with in the Phelps volume are given in e.g. Seater (1977 and 1978).

Originally, the relation between inflation and unemployment was formulated simply as

$$\pi_t = f(u_t) \tag{3.1}$$

where $\pi_t$ denotes inflation[3] in period t, $u_t$ denotes unemployment in period t, and f(.) is a convex, downward-sloping function for which there exists some $u^N$ such that $f(u^N) = 0$. This simple Phillips curve seemed to provide exactly the kind of information that the authorities needed for their policy decisions; given the relation (3.1), and given society's preference ordering in the $(\pi,u)$-space, we can choose the combination of inflation and unemployment that is optimal, i.e. that corresponds to the highest indifference curve $S(\pi_t, u_t)$ like in Figure 3.1:

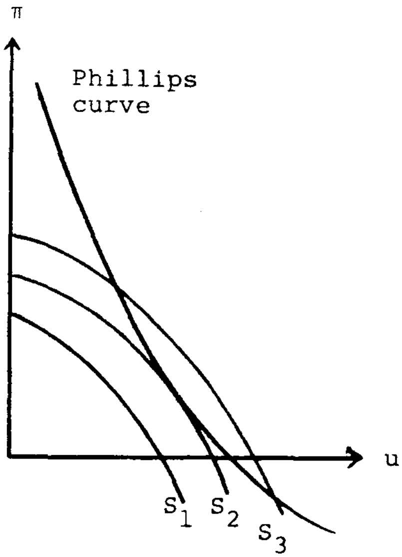


Figure 3.1

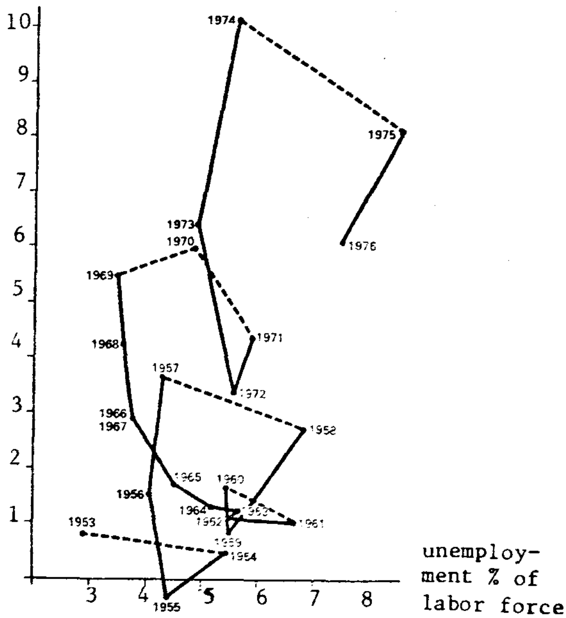


Figure 3.2: Inflation and unemployment in the U.S. 1953-1976 (Source: Economic Report to the President

---

[3] In fact, Phillips originally estimated *inflation in nominal wages* as a function of unemployment. It is however generally agreed that if some kind of Phillips relation holds, it will also hold (with the obvious change of sign and functional form) if unemployment is replaced by detrended real output, if money wage inflation is replaced by percentage changes in consumer prices, and so forth (cf. Lucas (1972 b, p. 50)). In the following, I will let $\pi$ denote the percentage change in consumer price index, but it could equally well be interpreted as the rate of change in any other relevant variable.

When inflation rates began to take off in the late 60's (cf. Figure 3.2) some economists questioned the appropriateness of a simple formula like (3.1), claiming that such a relation would hold only in the short run. Among the chief advocates of this view were Edmund Phelps (1967) and (1972), and Milton Friedman, who in his Presidential Address (1968) coined the expression "the natural rate" for the unique rate of unemployment that will prevail in a stable, long-run equilibrium in the neoclassical tradition, where demand equals supply for all goods (including leisure and time spent on job search) at the existing prices and wages. It seems quite reasonable to call the rate of "enemployment" that occurs in such a model "natural", and it is almost a truism to say that the model could be such that the natural rate of unemployment cannot be affected by monetary policy. The controversial question is, however, whether such a conclusion, drawn from such a model, is applicable to the economy of the real world, thereby implying that the long-run Phillips curve is vertical.

To formalize the view of the vertical Phillips curve, we take as our point of departure that the simple relation (3.1) does not tell the whole truth; inflationary expectations should be taken into account by adding a term to the formula:

$$\pi_t = f(u_t) + \alpha\pi_t^e \qquad (3.2)$$

where $\pi_t^e$ denotes expected inflation at time t.

Now there are two interpretations of the work "equilibrium" to deal with. The first one refers to the ordinary sense of the word: that demand equals supply at the present price system. The second meaning is what we will call *dynamic equilibrium*, which means that all variables follow stationary time series. We will not introduce the complications following from the stochastic approach yet, but illustrate the line of reasoning from a deterministic model. Thus, assume that unemployment is pegged to a constant u. If we define *dynamic equilibrium* as a state in which the rate of inflation is constant, and if we require for *consistency* that the expected rate of inflation equals the actual one, we have

$$\pi_t = \pi_t^e = \pi. \tag{3.3}$$

Substituting (3.3) into (3.2) we have a relation between real and nominal variables in a state of consistent, dynamic equilibrium which we call the long-run Phillips curve:

$$\pi = f(u) + \alpha\pi.$$

This shows that f(u) can be different from zero (i.e., u can be different from the so-called "natural rate" $u^N$) only if $\alpha \neq 1$. Rearranging, we obtain the expression

$$\pi = \frac{f(u)}{1-\alpha}$$

which shows that if $\alpha = 1$ and $u \neq u^N$, then no finite equilibrium inflation rate exists. Note that we have not yet specified how expectations are formed; the "natural rate property" of the above model holds *as an identity* as soon as we impose our definition of an equilibrium and our requirement for consistency, (3.3), on the model. Empirical investigations, however, aimed at finding out whether the natural rate hypothesis holds for the real economy, must of course specify how $\pi_t^e$ is formed in order to make estimates of the strategic parameter $\alpha$.[4]

We thus see that the answer to the second question on page 35 is somewhat ambiguous. As is evident from Figure 3.2, the relationship between inflation and unemployment is not as simple as the "naive" Phillips curve (3.1). However, there might exist a stable expectations-augmented Phillips curve (3.2) which, if $\alpha \neq 1$, implies that there exists a long-run

[4] Some commonly used specifications of $\pi_t^e$ are presented in Section 3.2 below.

trade-off between inflation and unemployment. *The third question* on page 35 concerned the role of economic policy, and one might be tempted to say that the answer to this question depends on the answer to the second: If $\alpha \neq 1$, then there is room for the authorities to perform employment policy, and if $\alpha = 1$ (i.e. if the natural rate hypothesis holds), then monetary policy, aimed at changing the rate of unemployment, is neither possible nor desirable. However, this is not true in general, and it is not true for our above model in particular.

To answer the question about the role of economic policy we first have to specify how economic policy is thought to work within the framework of the Phillips curve model. There are two fundamentally distinct ways of looking at economic policy, which are due to different interpretations of the Phillips curve. The earliest writers on the subject, including Phillips himself, and Lipsey (1960), seem to have assumed that the direction of causality runs from unemployment to inflation. The authorities were assumed to choose a time series of unemployment $\{\ldots, u_{t-1}, u_t, u_{t+1}, \ldots\}$, which could be more or less perfectly controlled by for example public works and different kinds of fiscal policy. The time series $\{u_t\}$ was transmitted through the Phillips model

$$\pi_t = f(u_t) + \alpha\pi_t^e$$

and, given the way expectations $\pi_t^e$ were formed, this resulted in an inflation time series $\{\ldots, \pi_{t-1}, \pi_t, \pi_{t+1}, \ldots\}$. The natural rate question in such a world was then: could the authorities choose a time series $\{u_t\}$ with an average value $E[u_t]$ different from the natural rate $u^N$ without resulting in an exploding (i.e. non-stationary[5]) time series $\{\pi_t\}$.

[5] We recall that for ARMA processes, non-stationarity means that the process veritably explodes if all the characteristic roots are within the unit circle; and it drifts away like the random walk if at least one root is on, and all the other roots are outside, the unit circle.

A more recent way of looking at economic policy, which is mostly associated with the Monetarist school, and which analyses the macro economy as an equilibrium system (in contrast with the earlier, "Keynesian" school), considers the opposite direction of causality. The authorities choose an inflation time series $\{\ldots, \pi_{t-1}, \pi_t, \pi_{t+1}, \ldots\}$, for example by means of the money supply. This series is transmitted through the economic system in a way which can be described by the Phillips curve, and an unemployment time series $\{\ldots, u_{t-1}, u_t, u_{t+1}, \ldots\}$ results. The natural rate question is then: Is there a stationary time series $\{\pi_t\}$ such that the corresponding series $\{u_t\}$ has a mean $E[u_t]$ different from the natural rate $u^N$?

There are thus two ways to model the conduct of economic policy, and although they are quite different on the conceptual level, they are very similar on the formal level. Given the model (3.2), we are interested in the mean of the time series $\{u_t\}$ and the stationarity properties of the time series $\{\pi_t\}$. On this level of aggregation, where the whole macroeconomic reality has been reduced to some simple properties of two time series connected by the Phillips relation, the direction of causality and the interpretation of the Phillips relation matters little, and the distinction between monetary and fiscal policy is of no importance. Thereby is not claimed that these questions are unimportant *per se*, but only that we have chosen to study the economy at such a level of abstraction that these things do not appear explicitly.

Another question which is of great importance to the role of economic policy, and which we will not pursue further, is the distinction between short and long run. The natural rate is a long-run concept, and even if the unemployment series $\{u_t\}$ cannot have a mean different from $u^N$, there might be room for economic policy in the short run. Edmund Phelps (1967) studies a model which displays natural rate properties, i.e. where inflation and unemployment are connected by

$$\pi_t = f(u_t) + \pi_t^e \tag{3.4}$$

and where expectations are formed adaptively[6] by

$$\pi_t^e - \pi_{t-1}^e = \delta(\pi_{t-1} - \pi_{t-1}^e) \tag{3.5}$$

and where we have the dynamic equilibrium and consistency condition

$$\pi_t = \pi_t^e = \pi. \tag{3.6}$$

For such a model the natural rate hypothesis obviously holds; as noted earlier, it follows from (3.4) and (3.6) regardless of how expectations are formed. However, the system (3.4) and (3.5) forms, together with a social welfare function with inflation and unemployment as arguments, a dynamic problem of optimal control. Phelps solves this control problem and characterizes the optimal paths $\{\pi_t\}$ and $\{u_t\}$; the intuitively reasonable result is that if society's time preference is high enough, it is acceptable to fight unemployment today even if such a policy results in higher inflation tomorrow. The answer to the last of our three questions on page 35 above is thus the following: *the fact that the natural rate holds for a model does not necessarily rule out the desirability of macroeconomic policy aimed at changing the unemployment rate.* Even if $u_t$ will be equal to $u^N$ in the long run, the Monetarist conclusion of non-intervention might be incorrect and rely on a static approach to policy problems that are essentially of a dynamic nature.

### 3.1.2 The Aggregate Supply Function

In the process of search for the micro-foundations of the Phillips trade-off, an analytical concept emerged which has

---

6 See Section 3.2.1.

become widely used during the last five years. It is the *aggregate supply function*, which relates real output to the price level by

$$Y_t = a + b(P_t - P_t^e) \tag{3.7}$$

where $Y_t$ is defined as the log of real output $y_t$, $P_t$ is the log of the price level $p_t$, and $P_t^e$ is an expectations variable. This particular function, while used also by earlier writers, has mainly been employed and advocated by Robert E. Lucas and his followers in the so-called rational expectations literature.[7]

There are several different ways of justifying a relation like (3.7). Lucas and Rapping (1969) for example, study a Fisherian model where workers allocate their time over two dates, choosing between labor supply and leisure today, and labor supply and leisure tomorrow. Assuming a utility function with consumption and working time as arguments, the rational utility maximizer then, according to Lucas and Rapping, solves the decision problem

$$\max U(c_1, c_2, \ell_1, \ell_2)$$

subject to

$$p_1 c_1 + \frac{p_2^e}{1+r} c_2 \leq A + w_1 \ell_1 + \frac{w_2^e}{1+r} \ell_2$$

where $c_1$ is consumption at time 1 and time 2, respectively, and $\ell_1$ and $\ell_2$ is labor supply at time 1 and time 2. $p_1$ and $w_1$ denote the price of consumption goods and the nominal wage at time 1, respectively; these two parameters are assumed to be known when the agent decides about consumption demand and

---

[7] Cf. Section 3.2.2.

labor supply at time 1. The future price and wage levels, $p_2$ and $w_2$, are however unknown at time 1, but the agent has formed expectations $p_2^e$ and $w_2^e$, respectively. Finally, r denotes the banks' lending and borrowing rate, and A denotes the agent's initial asset holdings. Solving the decision problem at time t = 1 leads to a labor supply function for period 1:

$$\ell_1^* = \ell_1^*(w_1, w_2^e, p_1, p_2^e, r). \tag{3.8}$$

Now we have a model for the supply of labor; by equalling this to the firms' demand for labor, which is obtained from the marginal productivity conditions of a CES production function, we obtain an expression for the supply of goods expressed as a function of among other things the price level $p_1$ and the price expectations $p_2^e$. This yields, after some fairly strong assumptions, a supply function like (3.7)

The micro-foundations of the aggregate supply function are not completely clear; especially the way the labor supply function (3.8) is transmitted through the goods markets is perhaps somewhat questionable.[8] There are however other ways to obtain the function (3.7) than the one just mentioned. The Lucas-Rapping model was a model with agents *speculating over time*; the possibility of a higher price or wage level tomorrow affects today's labor supply decision in a way that is reflected by the aggregate supply function. An equally important category of models is the one where agents *misperceive their real wages*.[9] If prices and wages increase by the same amount, the agents immediately observe the increase in their own (nominal) wages, but they obtain only a limited information of the increase in the general price level. Therefore they believe that their real wage has increased, this causing an

[8] The supply function will be critically examined in Chapters 6 and 7 below.

[9] This is the type of model implicit in M. Friedman (1968).

increase in the labor supply which, transmitted through the goods markets, gives rise to an aggregate supply function like (3.7).

In the above models the main emphasis was put on the behavior of the *labor force*. There are also related models where the *firms* misperceive the relative price of their output and speculate over markets, like in Laidler (1978). In this model the firm adjusts its output according to whether the price of its product is high in relation to the general price level:

$$y_{it} = f\left(\frac{p_{it}}{p^e_{it}}\right) \tag{3.9}$$

where $y_{it}$ is the output of firm i at time t, $p_{it}$ is the price observed by firm i in its particular market, and $p^e_{it}$ is firm i's perception of the general price level. Given some rather restrictive assumptions[10] we can sum over the firms to obtain

$$y_t = F\left(\frac{p_t}{p^e_t}\right)$$

which, if F is log-linear, yields

$$Y_t = a + b(P_t - P^e_t).$$

There are several important differences between the three types of models discussed above. *Firstly*, the uncertainty in the Lucas-Rapping model is borne by the workers, while the firms passively employ the labor supplied until the marginal productivity conditions are fulfilled. In Laidler's simple ad hoc model, the uncertainty is borne by the firms, and the labor supply does not appear at all in the model. In the intermediate, Friedman model, both interpretations

---

[10] For example, that all firms have identical reaction functions, and that all firms have the same perception of the general price level $p^e_t$.

are possible, Friedman's verbal description being general enough to comprise both labor and firms misperceiving their relative prices.

*Secondly*, the interpretation of the expectations term $p_t^e$ is different. In Lucas' and Rapping's Fisherian model $p_t^e$ obviously refers to the agents' (i.e. the workers') expectation of *future* prices, assuming today's general price level perfectly known. In Friedman's and Laidler's models, however, $p_t^e$ is interpreted as the agents' perception of *today's* general price level, which is not perfectly known. But at least for Laidler's model this dichotomy is perhaps not particularly important; it seems reasonable to conceive of a model where firms speculate over time instead of over markets and supply goods according to a formula like (3.9), where $p_{it}^e$ thus should be interpreted as an expectation of tomorrow's price level.

In contrast to the Phillips curve, all models from which the aggregate supply function has been derived are equilibrium models. It has been used exclusively in rational expectations models, and the direction of causality has been *from* nominal magnitudes *to* real magnitudes. Economic policy has thus been seen upon as a time series $\{\ldots, p_{t-1}, p_t, p_{t+1}, \ldots\}$ which can be controlled by e.g. the money supply. This time series is transmitted through the supply function (3.7) and, given a formula defining $p_t^e$, it produces a time series $\{\ldots, y_{t-1}, y_t, y_{t+1}, \ldots\}$. The natural rate hypothesis can then be stated in a way similar to the Phillips curve model: if no stationary time series $\{p_t\}$ exists such that the average value $\bar{Y}$ of the time series $\{Y_t\}$ is different from the "natural" level $y^N$, then the natural rate hypothesis is said to hold. Note however the stronger requirements in the supply equation model. While we for the Phillips curve only required the inflation rate $\pi_t$ to form a stationary series (which allows the *price level* to form a non-stationary series) we required for the supply equation that the price level be stationary. Therefore one should perhaps make a less strong definition for the latter case: if no stationary time series $\{\ldots, p_t, \ldots\}$ or $\{\ldots, (p_t - p_{t-1}), \ldots\}$ exist for which $\bar{y} \neq y^N$, then the natural rate hypothesis is said to hold.

We thus have two relations between real and nominal magnitudes: the Phillips curve (3.1) or (3.2), and the supply equation (3.7). Although they look quite different, the later stating that real magnitudes can be affected by the *rate of change* in the price level, they are not mutually exclusive. On the contrary all the models referred to above, from which (3.7) can be deduced, aimed at obtaining a Phillips-like relationship bewteen inflation and real income. Assume that the aggregate supply equation holds,

$$Y_t = a + b(P_t - P_t^e),$$

and that expectations are formed adaptively:

$$P_t^e - P_{t-1}^e = \delta(P_{t-1} - P_{t-1}^e).$$

Combining these two equations by using a Koyck transformation yields

$$Y_t = \delta a + b(P_t - P_{t-1}) + (1-\delta)Y_{t-1} \tag{3.10}$$

which is a Phillips-like relation between real output and the rate of inflation. A somewhat disturbing feature, though, is provided by the additional term $(1-\delta)Y_t$. What consequences will this have for the econometric estimation of the Phillips curve? Let us, by means of a production function, translate levels of real output into unemployment rates and write (3.10) as

$$\pi_t = \phi(u_t) + \psi(u_{t-1}) + \varepsilon_t \tag{3.11}$$

where $\varepsilon_t$ is a stochastic disturbance term.[11] Specifying the ordinary Phillips curve as

$$\pi_t = f(u_t) + \eta_t \tag{3.12}$$

we see that an estimation of (3.12) on (3.11) will probably give the impression of a significant Phillips relationship. However, since the disturbance term $\eta_t$ contains both $\varepsilon_t$ and $u_{t-1}$, the estimates of the parameters in f(.) will be biased if $u_t$ and $u_{t-1}$ are correlated. We thus see that a Phillips curve *can* be deduced from a micro model, namely the micro model that yields the supply equation, but that the estimation of the curve can cause problems due to the time series properties of the variable $u_t$. This shows that the Phillips curve and the aggregate supply function need not be mutually exclusive, but that the former is rather to be regarded as a reduced form of a larger system which contains the latter.

Of course, this does not mean that the Phillips curve and the supply equation *are* equivalent. We recall that the latter fomula relies on equilibrium micro models, and a Phillips curve derived from these is thus some kind of reduced form of an *equilibrium system*. However, the Phillips curve can also be interpreted in a "Keynesian" way, describing some features of a *disequilibrium system* with rigid prices. The underlying perception of the world is quite different in the two cases, even if the resulting functional forms are the same.

## 3.2 MODELS OF EXPECTATIONS

Economics, claiming to deal with real-world phenomena, must take uncertainty into account. The particular type of uncertainty which is related to the passage of time calls for

---

[11] Note that it is not at all self-evident that $\varepsilon_t$ should occur additively, or should have all the neat properties required for ordinary least squares estimation; we just assume this for simplicity.

a theory of expectations; in fact such a theory is an essential feature (and perhaps *the* essential feature) of macroeconomics. While this has been acknowledged throughout the history of macroeconomic theory, the "ex ante-ex post" analysis of the Stockholm School being an early representative of the role of expectations, explicit and operational models of *how expectations are formed* have hardly been used for more that the last few decades. In fact, only two such models exist, namely the *adaptive* and the *rational* expectations schemes. The former was the only one to be used during the 1950's and 1960's, while the latter has gained dominance during the 1970's.

### 3.2.1 Adaptive Expectations

The first time the concept of a *distributed lag* appears in the economic literature is in a paper by Irving Fisher (1925). In that paper, Fisher tries to explain the monthly volume of trade by the monthly change in the price level (i.e. a sort of Phillips relationship, with a reversed direction of causality). Noting that trade can hardly be affected immediately by the whole force of a price increase, but that the effect is rather likely to be spread out over some months, he obtained a substantial increase in correlation when he studied the relation not between trade volume and inflation $\pi_t$, but between trade volume and a variable $\pi_t^e$, defined as

$$\pi_t^e = \sum_{j=1}^{n} w_j \, \pi_{t-j}. \tag{3.13}$$

A distributed lag is generally defined as a linear relation between a variable Y and earlier values of a variable X:

$$Y_t = \Sigma w_j \, X_{t-j}. \tag{3.14}$$

In the particular case when the variable $Y_t$ can be interpreted as a forecast of the variable $X_t$, we talk of an adaptive expectation. Thus the distributed lag (3.13) defines an adaptive expectation of inflation.

The introduction of a distributed lag into the models serves two purposes: it increases the explanatory power (i.e. the $R^2$) of an estimation, and it introduces an element of dynamics into an otherwise static model. The first of these purposes, which thus is of an essentially econometric character, was the one that preoccupied Irving Fisher in his pioneering paper. In this context a problem arises: if the length n of the lag is large, the estimation of all the individual weights $w_i$ will cost a lot of degrees of freedom. If we can constrain the weights in some reasonable way we will make a gain in the estimation procedure. Irving Fisher thus assumed that the $w_i$:s were log-normally distributed, which meant that the entire lag structure was determined if only two parameters (the mean and the variance) of the distribution could be determined. In a later work Fisher (1930) used a linear lag structure instead, while Koyck (1954) and Almond (1965) developed methods for saving degrees of freedom by employing geometric and polynomial lag structures.[12]

The second aspect of distributed lags, that of providing a dynamic framework for the economic models, did not preoccupy the econometrically-oriented Fisher, who was mainly interested in improving the goodness of fit. The recognition that a formula like (3.13) or, more generally, (3.14), inserted into a suitable model, would yield certain dynamic patterns emerged in the mid-1930's with among others Erik Lundberg[13] and Jan Tinbergen,[14] and was formally investigated by Paul Samuleson (1939) in his famous "Multiplier-Accelerator"

---

[12] For extensive surveys of the history of distributed lags, see Nerlove (1958 and 1972) and Griliches (1967). A standard textbook on the subject is provided by Dhrymes (1971).

[13] Lundberg (1937).

[14] See e.g. Tinbergen (1935).

paper. These studies, however, relied on very simple, first-order lags, and it was furthermore not quite evident that the lags should be interpreted as expectations; they could equally well be the results of inertia in the production and investment processes, or some kind of adjustment costs.

The first macroeconomic study that explicitly models the formation of expectations and takes account of all the econometric and dynamic features of the distributed lags approach, is that of Cagan (1956). He developed a model for the holdings of real cash balances which in its simplest version assumes the desired real balances as a negative function of the expected inflation rate:

$$\log\left(\frac{m_t}{p_t}\right) = a + b\pi_t^e. \tag{3.15}$$

He then assumed that the expected rate of change in prices is revised each period of time in proportion to the difference between the actual rate of change and the rate of change that was expected:[15]

$$\pi_t^e - \pi_{t-1}^e = \delta(\pi_{t-1} - \pi_{t-1}^e). \quad 0 \le \delta \le 1 \tag{3.16}$$

Such a formula, which has already appeared once or twice in the above text, is called *adaptive* or error-learning: if my forecast was too low yesterday (i.e. $\pi_{t-1} - \pi_{t-1}^e$ is positive) I adapt by aiming a little higher today. The degree of adjustment to yesterday's error is determined by the coefficient $\delta$; a high $\delta$ means quick adjustment, and a correspondingly

---

[15] In fact, Cagan worked with a model in continuous time, while the above notation implies discrete time. The distinction is of no importance in the present context, but it should perhaps be mentioned that some of Cagan's stability results, which were obtained from the theory of differential equations (i.e. continuous time) do not hold for his empirical estimates, which are made on data collected at discrete intervals. On this question, cf. B Friedman (1975). For some other aspects of continuous versus discrete time in adaptive expectations models, see Burmeister and Turnovsky (1976).

greater risk for over-adjustment.[16] By successively lagging (3.16) and substituting, we obtain $\pi_t^e$ as an infinite distributed lag with geometrically declining weights:

$$\pi_t^e = \delta \sum_{j=o}^{\infty} (1-\delta)^j \pi_{t-1-j} \tag{3.17}$$

We thus see that every adaptive, error-learning scheme of the form (3.16) can be written as an infinite, geometric lag (3.17). Conversely, every lag structure of the form (3.17) can be expressed in the adaptive form (3.16). For more general lag structures

$$\pi_{t+\theta}^e = \sum_{j=1}^{\infty} w_j \pi_{t-j} \tag{3.18}$$

which are not necessarily geometric and the forecasts of which are not necessarily one-period-ahead forecasts, it is not generally true that they can be equivalently written according to a *simple first-order* adaption like (3.16). But the work "adaptive" does not exclusively refer to such a limited class of formulas; the adaption need not confine itself to the difference $(\pi_{t-1} - \pi_{t-1}^e)$, but could take into account differences of higher order $(\pi_{t-j} - \pi_{t-j}^e)$, the square of the differences, the rate of change of the differences, etc.[17] Therefore, I will use the term "adaptive" for all expectations of the form (3.18), even if they cannot be reduced to a simple formula like (3.16).

Now, combining the behavioral equation (3.15) with the expectations formula (3.16), and assuming that desired holdings of real balances are equal to actual holdings and that

---

[16] It is not certain that $\delta$ should be constant. In hyperinflations, for example, agents might become more sensitive to changes in the inflation rate, thereby adjusting quicker than in less dramatic periods. For a version of Cagan's model with a variable $\delta$, see Khan (1977).

[17] See Mincer (1969) for a discussion of some more complicated adaptions.

nominal money supply $m_t$ is independent of $p_t$, we obtain a difference equation in $p_t$ which might be stable or unstable depending on the parameters $\delta$ and b.[18] Cagan's model thus not only provides a framework which has a "reasonable" resemblance to reality, but also casts the economy in a simple, dynamic setting, thereby making it possible to arrive at conclusions concerning stability, oscillations, etc.

At the same time as Cagan used the geometric lag formula (3.16) or (3.17), another Chicago economist employed it for the construction of time series of income, namely Milton Friedman in his "A Theory of the Consumption Function". While Cagan's $\pi_t^e$, however, is a genuine expectations variable, Friedman's permanent income $y_p^*(t)$ does not *exactly* correspond to the agent's expectation of tomorrow's income. From one point of view, $y_p^*(t)$ can be regarded as the expectation of a probability distribution, which makes it look like a genuine expectations variable, while from another point of view it should take into account not only the probability distribution of tomorrow's income but also income over the future lifetime of the individual. When discussing the exact interpretation of $y_p^*$ Friedman takes a fairly pragmatic view of these problems,[19] and we should perhaps not go too deep into them.

Now, the geometrically distributed forecasts of Cagan and Friedman gained a wide acceptance during the 1950's and were used in numerous studies of economic phenomena. Apart from the simplicity and the attractive error-learning feature that lies in formula (3.16), they were used only on an *ad hoc* basis without any questioning of whether, or why, the adaptive expectations were justified on more statistical grounds. This led Muth (1960) to ask the following question:

---

18 With our formulation of the model (in discrete time), the equation becomes quite tedious, while Cagan's assumption of continuous time yields a fairly simple differential equation.

19 Cf. Friedman, *op.cit.*, chapters III and VI.

Assume we have a time series $\{\ldots, y_{t-1}, y_t, y_{t+1}, \ldots\}$ of which we make an adaptive forecast

$$y_t^e - y_{t-1}^e = \delta(y_{t-1} - y_{t-1}^e). \quad 0 \le \delta \le 1 \qquad (3.19)$$

For what kind of time series $\{y_t\}$ is the forecast $y_t^e$ appropriate, i.e. for which time series is $y_t^e$ an "optimal" forecast? To answer this, he first defined an optimal forecast as a forecast $y_t^e$, given by (3.19), such that the mean square error

$$V \equiv E[(y_t - y_t^e)^2]$$

be minimized.[20] He then wrote the stochastic process $\{y_t\}$ on the general moving-average form

$$y_t = \varepsilon_t + \sum_{j=1}^{\infty} \beta_j \varepsilon_{t-j} \qquad (3.20)$$

By assuming that $E[y_t] = 0$, i.e. that there is no intercept in the process (3.20), we see that *for this particular case* the optimal predictor $y_t^e$ can coincide with the conditional expectation of $y_t$. The question is then: For which process (3.20), i.e. for what parameters $\beta_j$, does (3.19) provide a forecast such that

$$y_t^e = E[y_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots].$$

Muth shows that for this to hold, $\{y_t\}$ must be a *modified random walk* of the form

$$y_t = \varepsilon_t + \delta\varepsilon_{t-1} + \delta\varepsilon_{t-2} + \delta\varepsilon_{t-3} + \ldots \qquad (3.21)$$

It is in fact possible to generalize the result a bit further. The formula (3.19) assumes that we are only interested in forecasts of the time period immediately ahead. Assume instead that we have a *forecasting span* of $\theta$ periods, i.e.

[20] This criterion can of course be questioned, but it is the standard criterion for optimality in econometrics and forecasting literature.

that we form the expectation

$$y^e_{t+\theta} - y^e_{t+\theta-1} = \delta_\theta(y_{t-1} - y^e_{t-1}). \tag{3.19'}$$

Muth demonstrates that this forecasting formula is optimal for a process of the form

$$y_t = \varepsilon_t + \delta_\theta\varepsilon_{t-1} + \delta_\theta\varepsilon_{t-2} + \delta_\theta\varepsilon_{t-3} + \ldots$$

We can thus drop the subscript of the coefficient $_\theta$ and conclude that if the time series to be forecasted follows the pattern (3.21), a forecast of the form (3.19') is optimal *regardless of the forecasting horizon* $\theta$. In other words, for the modified random walk (3.21) we have

$${}_{t-1}y^e_t = {}_{t-1}y^e_{t+1} = {}_{t-1}y^e_{t+2} = \ldots = {}_{t-1}y^e_{t+\theta}$$

(where the first subscript of ${}_{t-1}y^e_{t+\theta}$ refers to the time at which the forecast is formed).

We thus have the result that if (3.19) is to be an optimal forecast of a process that can be written as (3.20), and if

$$y^e_t = E[y_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots], \tag{3.22}$$

then the $\beta_j$ coefficients must all be equal to $\delta$. But what if the process is such that the conditional expectation cannot be expressed as a weighted sum of earlier realizations, i.e. what if (3.22) does not hold? Will (3.19) be optimal for these, too, or is there another set of adaptive weights $w_i$ that minimizes the mean square error?

Muth studied one such process, namely the process where $y_t$ is assumed to consist of two separate parts, a "permanent" part $x_t$ and a "transitory" part $\eta_t$:

$$y_t = x_t + \eta_t. \tag{3.23}$$

Assume that the transitory variable is white noise, while the permanent variable follows an ordinary random walk,

$$x_t = x_{t-1} + \varepsilon_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \dots$$

In this case the conditional expectation of $y_t$ is

$$E[y_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \eta_{t-1}, \eta_{t-2}, \dots] = E[x_t | x_{t-1}] = \\ = x_{t-1}$$

However, since both the permanent part $x_t$ and its transitory counterpart $\eta_t$ are unobservable, the only observable variable being their sum $y_t$, we might have difficulties in forming the expectation $E[y_t | \dots]$. What we can easily do instead is to form a forecast $y_t^e$, not necessarily equal to $E[y_t | \dots]$, as a weighted sum of earlier observations of y:

$$y_t^e = \sum_{j=1}^{\infty} w_j \, y_{t-i} \tag{3.24}$$

A forecast according to (3.24) is thus not always "rational",[21] but we can nevertheless ask what the adaptive weights $w_i$ should look like if we want to obtain an $y_t^e$ on the form (3.24) which minimizes the mean square error V. It then turns out that the $w_j$:s will be of the form

$$w_j = \delta(1-\delta)^{j-1} \qquad i = 1, 2, 3, \dots, \tag{3.25}$$

which means that the forecasts will be geometric like those in (3.19), although the relation between $\delta$ and the parameters of the process $\{y_t\}$ (i.e. the variances $\sigma_\varepsilon^2$ and $\sigma_\eta^2$ and the covariance $\sigma_{\varepsilon\eta}^2$) will be quite complicated.

Muth's results were not entirely new; similar forecasting schemes had earlier been analyzed in the statistical li-

[21] Cf. Section 3.2.1 below.

terature[22] by means of general, but also quite difficult, spectral methods hardly accessible to the economists working with expectations in macro models. Muth's contribution gave a much simpler approach to the theory of optimal forecasting, and was followed by a number of studies employing similar methods. Bailey (1960 and 1971) studied the general problem of forecasting a process $\{y_t\}$ subject both to disturbances $x_t$ and to errors of observation $\eta_t$:

$$y_t = x_t + \eta_t.$$

This is similar to Muth's equation (3.23), but while Muth had assumed $x_t$ to follow a random walk process and $\eta_t$ to be simply white noise, Bailey assumed $x_t$ and $\eta_t$ to follow general, stationary processes

$$x_t = \sum_{j=1}^{n} \alpha_j x_{t-j} + \varepsilon_t$$

$$\eta_t = \sum_{j=1}^{m} \beta_j \eta_{t-j} + v_t.$$

The problem of finding optimal weights to the formula

$$y_t^e = \sum_{j=1}^{\infty} w_j y_{t-j}$$

is not as easy as in Muth's case and the $w_j$:s are not in general of the simple geometric type. Furthermore, the independence of the forecasting horizon $\theta$ is not preserved.

Contemporary with Bailey's work were some papers by Nerlove and associates[23] dealing with a slightly different problem. In Muth's and Bailey's papers the problem was to forecast the observable variable $y_t$ as a weighted sum of its own earlier values. The parallel to Milton Friedman's permanent income hypothesis, however, requires rather the *unobserv-*

---

[22] Cf. Yaglom (1962) for an account of these results.

[23] See Nerlove (1967).

*able* permanent income $x_t$ to be forecasted as a weighted sum of earlier values of the observable total income $y_t$:

$$x_t^e = \sum_{j=o}^{\infty} w_j \, y_{t-j}. \tag{3.26}$$

In Muth's model the two problems are somewhat similar; we have that

$$E[y_t | \ldots] = x_{t-1} = E[x_t | \ldots],$$

since the permanent income $x_t$ was assumed by Muth to follow a random walk process $x_t = x_{t-1} + \varepsilon_t$. A "good" forecast of $y_t$ is therefore also a "good" forecast of $x_t$, given the particular structure of the model. However, the two models are not entirely equivalent, and the optimal $w_j$ weights for forming $y_t^e$ are in general not identical with the $w_j$ weights for forming $x_t^e$.

Nerlove sets out to find the w vector of (3.26) which is optimal in the sense of minimizing the mean square error

$$V \equiv E[(x_t - x_t^e)^2].$$

The problem is quite complicated for more general processes $\{x_t\}$ and $\{y_t\}$, but for simple processes some elegant results are to be obtained. Assume for example that total income is the sum of permanent and transitory incomes,

$$y_t = x_t + \eta_t,$$

and that permanent income follows a first-order autoregressive process

$$x_t = \psi x_{t-1} + \varepsilon_t, \qquad 0 \leq \psi < 1$$

the processes $\{\eta_t\}$ and $\{\varepsilon_t\}$ being white noise with variances

$\sigma_\eta^2$ and $\sigma_\varepsilon^2$, respectively. It can then be shown that the optimal prediction is given by

$$x_t^e = \frac{\psi-\beta}{\psi} \sum_{j=o}^{\infty} \beta^j \, y_{t-j} \tag{3.27}$$

where

$$\beta \equiv \frac{(1 + \sigma_\varepsilon^2/\sigma_\eta^2 + \psi^2) - \sqrt{(1 + \sigma_\varepsilon^2/\sigma_\eta^2 + \psi^2)^2 - 4\psi^2}}{2\psi} .$$

Letting $\psi$ approach unity from below[24] we see that the weights in (3.27) become quite similar to the weights in Muth's problem (3.25) with $\psi \equiv (1-\beta)$.[25]

We have thus seen that there has been three lines of development in the theory of adaptive expectations. The first line (which was first also in chronological order) concerned only estimation procedures; the problem was to find a distributed lag which explained as much as possible of the variations in some dependent variable. This task was performed by applying a number of lag structures, the most prominent ones being Irving Fisher's linear, Koyck's and Cagan's geometric, and Almon's polynomial structure. The second line of research concerned the analysis of dynamic adjustments; the problem was to obtain a difference equation which was simple enough for being studied analytically, and this task was performed mainly by applying the geometric lag. The last line, which was not developed until during the 1960's, paid tribute to the economists' taste for optimizing behavior; the problem was to find out whether the adaptive expectations

---

[24] $\psi$ must be $< 1$, since the derivation of formula (3.27) holds only for stationary processes.

[25] Note however the conceptual differences; (3.25) refers to the prediction of $y_t$ by means of $y_{t-1}$, $y_{t-2}$ etc., while (3.27) refers to the prediction of $x_t$ by means of $y_t$ (the value of which is already available), $y_{t-1}$, $y_{t-2}$ etc.

schemes minimized the mean square error of some stochastic process. This analysis was mainly concentrated upon the geometric lag structure, although some more general results were obtained.

Of course one can always criticize the idea of modelling expectations as a sum of earlier values of the variable to be forecasted. Although it proved both analytically and empirically fruitful, the idea does not correspond very well to how expectations are formed in reality (as can be proved by a moment's introspection). It has already been mentioned that for some time series, the conditional expectation cannot be expressed as a weighted sum of earlier realizations. The question thus naturally arises: Why should we care about that weighted sum at all? Why not make a forecast according to

$$y_t^e = E[y_t | H_{t-1}],$$

where $H_{t-1}$ stands for the past history of the model generating the variable $y_t$, instead? This way of modelling expectations, which turned out to be as operational, both on the empirical and the analytical level, as any alternative scheme - thereby fulfilling the qualifications for a good model - came into use at the beginning of the 70's, drawing most of the attention from the adaptive expectations scheme.

### 3.2.2 Rational Expectations

In his 1960 paper John F. Muth, as we have seen, discussed the question of which stochastic processes can be optimally forecasted by an adaptive forecasting formula with geometrically declining weights. Note however that he regarded the stochastic process as exogeneously given; he did not ask whether some kind of economic model would generate such a particular process. In a subsequent paper[26] he went one step

[26] Muth (1961).

further and studied a simple economic equilibrium model with an expectations term and a stochastic disturbance in the supply equation. Assuming that the agents formed optimal forecasts of future prices he solved the model for the price variable and studied in particular whether the resulting time series $\{\ldots, p_t, \ldots\}$ was such that its conditional expectation, $E[p_t | H_{t-1}]$, could be expressed as a sum of earlier p values with geometrically declining weights. This paper, being the first to explicitly introduce the concept of "rational expectations" into economics,[27] was disregarded for more than a decade, but after its rediscovery in the early 1970's it has had a great impact on macroeconomic theory. It thus seems appropriate to present its results in some detail.

Muth considered a model of market equilibrium in which the consumers observe the price ruling in the market and form their demand according to an ordinary, downward-sloping demand curve. The producers, on the other hand, are subject to a one-period production lag; at time t-1 they have to decide how much to supply at time t. When the decision is taken they do not know tomorrow's price $p_t$, so therefore they make their decision by considering the expected price, $p_t^e$. Furthermore, the production process is subject to a stochastic disturbance, so the actual supply at time t differs from the planned supply by an error term $e_t$. The model thus reads[28]

$$\left.\begin{array}{ll} \text{Demand:} & D_t = -\beta\, p_t \\ \text{Supply:} & S_t = \psi\, p_t^e + e_t \\ \text{Market equilibrium:} & D_t = S_t \end{array}\right\} \quad (3.28)$$

---

[27] John Rutledge (1974) has traced the concept back to Keynes, Marshall, and even to John Locke. This is true, in the sense that these mainly verbal writers have assumed a certain degree of rationality or consistency of their agents, which can perhaps be interpreted ex post in terms of Muth's rational expectations.

[28] All variables are defined as deviations from their equilibrium values. This assumption does not restrict the conclusions, but leaves out a few constants in the expressions below.

The error term $e_t$ is unknown at the time the production decisions are made, but it is known - and relevant - at the time the commodity is purchased in the market. Now, Muth's main point is that information is scarce, and the economic system does not therefore waste it. The agents thus make as good forecasts as possible, not necessarily confining themselves to any fixed adaptive scheme like the ones presented in the preceding section. The "best" forecast one can make in terms of mean square error is the mathematical expectation[29] *conditional upon the particular structure of the model* and upon the previous disturbances; thus Muth postulates that the expectations variable is given by

$$p_t^e = E[p_t | e_{t-1}, e_{t-2}, \ldots].$$

Solving the above model for $p_t$ and taking expectations yields

$$p_t^e = -\frac{1}{\beta+\psi} E[e_t | e_{t-1}, e_{t-2}, \ldots]$$

which, by assuming that $e_t$ follows an arbitrary moving-average process

$$e_t = \sum_{j=1}^{\infty} \alpha_j \varepsilon_{t-j} + \varepsilon_t, \tag{3.29}$$

can be written on the form

$$p_t^e = -\frac{1}{\beta+\psi} \sum_{j=1}^{\infty} \alpha_j \varepsilon_{t-j}. \tag{3.30}$$

We are now interested in writing $p_t^e$ on the adaptive form

$$p_t^e = \sum_{j=1}^{\infty} w_j p_{t-j} \tag{3.31}$$

which is a simple task since we know the moving-average co-

---

[29] For a simple proof that the expectations operator $x_t^e = E[x_t | \ldots]$ minimizes the mean square error $V \equiv E[(x_t - x_t^e)^2]$, see Nelson (1973), p. 143 f.

efficients of (3.30). The $w_j$ weights can thus be computed as functions of the parameters in the system (3.28) and the $\alpha_j$:s in the $\{e_t\}$ process. The results can be illustrated by two examples. Firstly, assume that the $e_t$ disturbances are independently distributed and thus follow a white noise process

$$e_r = \varepsilon_t, \tag{3.29'}$$

i.e. $\alpha_j = 0$, $i = 1, 2.$ ... Computing the $w_j$ weights in (3.31) then gives

$$p_t^e = 0,$$

i.e. $w_j = 0$, $j = 1, 2, \ldots$

For a second example, we assume that $e_t$ follows a random walk, which means that $\alpha_j = 1$ for all j:

$$e_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots \tag{3.29"}$$

Such a process yields optimal forecasting weights according to

$$p_t^e = \frac{\beta}{\psi} \sum_{j=1}^{\infty} \left(\frac{\psi}{\beta+\psi}\right)^j p_{t-j} \tag{3.32}$$

This model thus generates prices that follow a process which can be optimally predicted by a geometric lag structure, although the lag distribution is slightly more complicated than Cagan's $w_j = \delta(1-\delta)^j$.

Muth did not apply his concept of rational expectations to macroeconomics; instead he studied cobweb theorems and showed that his model, where the agents are *not* systematically misjudging the prices, generates cycles which can more satisfactorily explain observed price movements in hog and cattle markets than can the older adaptive-expectations models. This emphasis of his is probably one explanation to the fact that his approach was neglected in macroeconomic analysis for more than a decade. Another reason for this is probably the ambiguity of the concept of rational expectations itself. The rationality implies that the agent somehow realizes that his expectation affects prices, and that he, with his insight, forms an "optimal" expectation. But optimal could mean several things; one could for example talk about the *optimal adaptive expectation*, i.e. the set of weights $w_j$ for which the expectation

$$p_t^e = \sum_{j=1}^{n} w_j \, p_{t-j} \tag{3.33}$$

minimizes the mean square error. One could also disregard the restriction that the adaptive formula (3.33) imposes on the forecast and define as optimal the forecast $p_t^e$ which, *without any restrictions*, yields the smallest mean square error. This means that $p_t^e$ is given by the "true" rational expectation

$$p_t^e = E[p_t | H_{t-1}], \tag{3.34}$$

where $H_{t-1}$ denotes past history of the model through time t-1.

In Muth's model there is no contradiction between the two formulas. He constructed it in such a way that the rational expectation (3.34) can always be expressed as a weighted sum of earlier observations (3.33), if necessary with $n = \infty$. This diffused the distinction between the two approaches and Nelson (1975) has suggested that it might be one of the reasons to why Muth's paper was not really regarded as a critique of the adaptive expectations approach. The distinction between the two approaches is clearcut, however. Let us return to the equilibrium model (3.28) but let us assume, in addition, that there are two stochastic disturbances instead of one, i.e. that the demand function reads

$$D_t = -\beta\, p_t + v_t. \tag{3.28'}$$

If $e_t$ and $v_t$ disturbances are independently distributed, but autocorrelated, time series, the conditional expectation of $p_t$ cannot be expressed as a weighted sum of earlier prices. For stationary time series, $p_t$ will follow a stationary process too, and according to Wold's decomposition theorem $p_t$ can then be expressed in terms of an autoregressive process

$$p_t = \sum_{j=1}^{\infty} w_j\, p_{t-j} + \eta_t. \tag{3.35}$$

A predictor $\hat{p}_t$ may therefore be written in the form

$$\hat{p}_t = \sum_{j=1}^{\infty} w_j\, p_{t-j}$$

which is an adaptive expectation. The best predictor of $p_t$ is however the conditional expectation $p_t^* \equiv E[p_t | e_{t-1}, e_{t-2}, \ldots, v_{t-1}, v_{t-2}, \ldots]$; it can be shown that, for all time series $\{e_t\}$ and $\{v_t\}$,

$$E[(p_t-\hat{p}_t)^2] \geq E[(p_t-p_t)^2],$$

where the equality holds only if $\{e_t\}$ and $\{v_t\}$ are white noise processes.[30] The reason for this is that when moulding the model (3.28-3.28') into the autoregressive process (3.35), some information is lost, and the stochastic disturbances $\eta_t$, which are merely statistical artifacts constructed out of earlier realizations of the $\{e_t\}$ and $\{v_t\}$ processes, have

[30] This is demonstrated in Nelson (1975).

such a large variance that forecasts building on (3.35) necessarily become less precise than "rational" forecasts.

When Muth's paper was rediscovered in the early 1970's, it served as a stimulus to substantive research into established models recast in the rational expectations framework. For example, Sargent and Wallace (1973) and Black (1974) reexamined the demand for cash balances during hyperinflations (cf. page 51 above) with Cagan's (1956) adaptive expectations substituted by rational expectations, while Brock (1974) analyzed a growth model in the spirit of Sidrauski (1967) where inflationary expectations are formed rationally.[31]

The papers that are most relevant for us, however, are those explicitly dealing with the natural rate hypothesis in stochastic macro models. The key paper was published by Lucas (1972 a) and analyses an overlapping-generations model where the individuals live for two periods, working during the first and saving money for their second, retirement, period. The government issues fiat money which performs one function only: to carry over individuals' wealth from the first period to the second. Now, the monetary phenomena that occur are due to uncertainty and lack of information, modelled by assuming that the individuals are divided between two markets with no communications between them. There are two kinds of stochastic disturbance, one real and one nominal. The former is accomplished by assuming that at each time period, the distribution of individuals between the two markets is a stochastic variable: a fraction $\theta/2$ of the young generation goes to the first market and a fraction $1 - \theta/2$ goes to the second, the parameter $\theta$ being drawn from a lottery with a known probability distribution. The old generation,

[31] All these models are deterministic, and are therefore not appropriate to analyse the particular macroeconomic phenomena that are due to uncertainty. Determinstic models with rational expectations are often called *perfect foresight models*.

which carries over money balances from the previous period and buys goods from the young, is divided equally between the markets. This means that real wages can and, in general, will differ between the two markets. The nominal disturbance, on the other hand, is accomplished by the authorities monetary supply rule, which contains a stochastic element. If the nominal money supply at time t, given to the individuals of the younger generation and divided equally between the two submarkets, is denoted by $m_t$, the money supply at time t+1 is given by

$$m_{t+1} = x \cdot m_t \tag{3.36}$$

where x is a random variable with known probability distribution. A first-generation individual has to decide how much to save for his old age. He is thus facing the following decision problem under uncertainty:

$$\max_{c,n,\lambda\geq o} U_t(c_t,n_t) + E[U_{t+1}(c_{t+1})] \tag{3.37}$$

subject to

$$p_t(n_t-c_t) - \lambda \geq 0$$

$$p_{t+1}\, c_{t+1} = \lambda,$$

where $n_t$ is the individual's labor supply in period t, $p_t$ is the price level (labor and consumption goods having the same price), and $\lambda$ is the amount of money balances carried over by that particular individual from period t to t+1. When solving this problem, he cannot quite distinguish between real and nominal disturbances, even if he has rational expectations in the sense of knowing the probability distributions of $\theta$ and x. Thus, observing a high price, or wage, level $p_t$ could mean *either* that the agent of the younger generation has happened to be allocated to a market with comparatively few persons of productive age, *or* that the total money supply has been large at that particular time. Without knowing which, he determines his labor supply $n_t$ and first-period consumption demand $c_t$ by solving (3.37). As a retired person, in the second period of his life, the decision problem is simple: money cannot be inherited, so old-age consumption is identically given by $c_{t+1} = \lambda/p_{t+1}$.

From this model, where rational agents cannot distinguish between real and nominal disturbances, a number of interesting properties can be derived. *First*, it will give rise to Phillips-like fluctuations in economic activity: periods with high prices will also be characterized by high output (i.e. high labor supply $n_t$), and it will also hold that a regression of the *rate* of price increases on the level of output will yield significant coefficients. The reason for this is that the individuals will misinterpret an inflationary price increase as a rise in real wages, thereby tending to increase their labor supply. *Second*, all the movements along the Phillips curve will be of a purely stochastic nature, which means that any systematic attempts to move it along the curve will be futile. In other words: there exists no money supply rule like (3.36) which can affect the average labor supply $E[n_t]$ or the average level of real output $E[y_t]$. *A third result* is that among all conceivable money supply rules, one is pareto-optimal, namely the one minimizing the nominal disturbances by implementing a constant, pre-announced growth rate with no stochastic components in the money supply.[32]

---

[32] The whole setting of the model is thus an attempt to formalize the ideas expressed by Milton Friedman (1968), where the agents are misperceiving their real wages, and to demonstrate that Phillips-like patterns emerge even if the agents are not displaying any money illusion but are perfectly rational in the sense of knowing the probability distribution of the variables $\theta$ and x. Also the result concerning the optimality of a stable money supply rule refers to an idea by Friedman, namely the so-called k-percent rule. Cf. Friedman (1960).

Lucas' model might seem a bit peculiar in the institutional setting described above, but it nevertheless served its purpose of giving a consistent picture of macroeconomic phenomena occurring in a world with rational agents having less than perfect knowledge. It was followed by a number of papers on a less rigorous level with more or less simple *ad hoc* models, oriented towards policy conclusions and empirical testing rather than towards a coherent general equilibrium framework. The first of these was also by Lucas (1972 b), and displays an aggregate supply equation

$$Y_t = a + b(P_t - P_t^e), \tag{3.38}$$

where $Y_t$ denotes the log of real, aggregate output, $P_t$ is the log of the price level, and $P_t^e$ is a price expectations term. This is substituted into an aggregate demand schedule, or rather a price identity

$$Y_t + P_t = X_t,$$

where $X_t$ denotes the log of nominal, aggregate demand. Assuming that the authorities can affect $X_t$ by some monetary policy rule, such as for example

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \varepsilon_t \tag{3.40}$$

the system forms a closed macro model which, with the assumption of rational expectations,

$$P_t^e = E[P_t | H_{t-1}],$$

turns out to have the property that no monetary policy rule, i.e. no parameters $\alpha_1$ and $\alpha_2$, can affect the real variable $Y_t$. A model in the same spirit, but somewhat more complicated, is given in Lucas (1973). There the aggregate *ad hoc* approach is taken from the 1972(b) paper, while the combination of real and nominal disturbances, and the agents' inability to distinguish between the two, is taken from the 1972(a) paper. The conclusion is the same, namely that monetary policy cannot systematically affect real output, and the model is supported by some empirical tests of time series from different countries.[33]

Now these early ad hoc models, being of a rather simple structure which for example excluded the financial sector of the economy, were soon followed by more "complete" macro models. For example, a typical second-generation rational expectations model includes first an aggregate demand, or IS, schedule:[34]

$$Y_t = c_1 + c_2[r_t - ({}_{t-1}P^e_{t+1} - {}_{t-1}P^e_t)] + u_t,$$

where $r_t$ denotes the nominal interest rate and ${}_{t-1}P^e_{t+1}$ denotes the (log of the) price level expected at time t-1 to prevail at time t+1. That is, $({}_{t-1}P^e_{t+1} - {}_{t-1}P^e_t)$ is the rate of inflation expected at time t-1 to prevail between t and t+1, i.e. the variable in square brackets relates to the expected *real rate of interest*. We add a portfolio balance, or LM, schedule

$$m^d_t = P_t + d_1 Y_t + d_2 r_t + v_t \qquad (3.41)$$

where $m^d_t$ stands for the log of the demand for nominal money balances. We can now formulate the authorities' money supply rule

$$m^s_t = g\ \theta_{t-1} + \varepsilon_t \qquad (3.42)$$

where $\theta_{t-1}$ represents the set of current and past values of all of the endogeneous and exogenous variables through period

---

33 An erratum to these tests is given in Lucas (1976 b).

34 This draws heavily on Sargent (1973) and Sargent and Wallace (1975), although the equations are somewhat simplified.

t-1 and g is a vector of parameters conformable to $\theta_{t-1}$. For the particular case of a purely autoregressive pattern in the money supply, we would set for instance

$$g = (\alpha_1, \alpha_2, \ldots, \alpha_n)$$

and

$$\theta_{t-1} = (m^s_{t-1}, m^s_{t-2}, \ldots, m^s_{t-n})^T.$$

Now all this, together with the equilibrium condition $m^s_t = m^d_t$ constitutes an ordinary textbook example of a "Keynesian" economy, with a liquidity preference schedule defined by (3.41). For such a model we are used to deducing optimal stabilization rules (3.42). If we however add the aggregate supply function

$$Y_t = a + b(P_t - {}_{t-1}P^e_t) \qquad (3.43)$$

together with the assumption of rational expectations,

$$P^e_t = E[P_t | \theta_{t-1}], \qquad (3.44)$$

the system turns out to become completely blocked, the variables $E[Y_t]$ and $E[Y_t | \theta_{t-1}]$ being independent of the control vector g in (3.42). The Monetarist conjecture of a "natural rate of unemployment" thus holds even for a model with liquidity preference - if (3.43) and (3.44) are added to it.

This type of model was extended by Barro (1976) to include features from Lucas (1972 a) and (1973): the distribution of agents to different markets with no information flows between them, and the inability to distinguish between real and nominal disturbances. Phelps and Taylor (1977) and Fisher (1977) modified it by introducing the assumption that some agents enter multi-period contracts, thereby committing themselves over several periods, while the monetary authorities can adjust their actions every period. The authorities thereby have an "information advantage" which allows them to affect the real variables - contrary to the Monetarist conjecture. Some scope for monetary policy exists also during the "transition to rational expectations", i.e. the time before the agents have collected enough data to estimate the whole system and to compute the conditional expectation (3.44), as is shown by Taylor (1976) and B. Friedman (1979).

And this brings us finally to a basic question in the rational expectations literature: Do the agents really form rational expectations? If even professional economists disagree about the true structure of the economic system, how can then the local retail store-keeper or the unemployed job-seeker have enough information to form "true" mathematical expectations? In Shiller's critical review (1978) this is put forward as one of the main weaknesses in the concept of rational expectations, while Barro and Fisher in their survey article dismiss that type of criticism as "reminiscent of criticisms of microeconomic theory on grounds that most consumers have never seen a Lagrangean, and it is equally beside the point".[35] In fact, Muth (1961) is quite clear on the question that individual agents do *not* necessarily form optimal forecasts, but that the market, as an aggregate, somehow works *as if* it can be described by a set of equations (3.28) with $p_t^e = E[p_t | H_{t-1}]$. This reasoning has a certain intuitive appeal, but it still remains to be shown how not perfectly rational agents can be aggregated so that the market works as if "it" were rational. Any model demonstrating how such aggregate behavior would occur has not yet been presented, and the assumption of rational expectations does not seem particularly more (or less) realistic than that of its predecessor, the adaptive expectations. Lacking sufficiently good data on people's actual

---

[35] Barro and Fisher (1976, p. 113).

expectations,[36] testing one expectations formula or another is really a joint test of the particular model *and* the particular expectations mechanism; it is thus hardly surprising that decisive test results do not yet exist. The two alternative expectations mechanisms can both be criticized on reasonable grounds. On the other hand - since those are the only ones existing, they both deserve to be studied, and the economic policy implications of both of them should be analyzed.

[36] Some tests of the rationality of economic forecasts have been performed by e.g. McNees (1978) and B. Friedman (1978).

# 4. ADAPTIVE EXPECTATIONS: THE OPTIMAL WEIGHT SUM

The Monetarist conjecture that money is neutral, i.e. that the long-run Phillips curve is vertical, implies that the "money illusion parameter" $\alpha$ in the Phillips curve,

$$\pi_t = f(u_t) + \alpha \, \pi_t^e, \tag{4.1}$$

is equal to unity. Substantial effort has been spent on estimations aimed at accepting or rejecting the hypothesis that $\alpha \neq 1$. A problem has been that the strategic variable $\pi_t^e$ is unobservable, thereby forcing the econometricians to rely on an observable proxy. The most widely employed proxy has been the adaptive expectations scheme

$$\pi_t^e = \sum_{j=1}^{n} w_j \, \pi_{t-j}. \tag{4.2}$$

Two constraints have generally been imposed on the $w_j$ weights, namely

$$w_j \geq 0, \text{ all } j,$$

and

$$\Sigma w_j = 1.$$

The estimates thereby obtained of the money illusion parameter $\alpha$ have yielded somewhat inconclusive results. $\alpha$ has usually been below, but not always significantly below, unity, and it seems to have been increasing over the last years, since more recent studies yield higher values than do earlier ones.[1]

In order to make as good judgements as possible about the reliability of the $\alpha$ estimates, it is important to know the properties of the expectations formula employed. In particular, it is important to see whether the constraints imposed on the $w_j$ weights might have had any effects on the estimates of $\alpha$, and in that case what effects.

The first of the two constraints, that of non-negativity in the $w_j$ weights, is imposed because together with the second it makes it possible to regard the weights as probabilities, and to treat the distribution of weights as a probability distribution. Thereby one can define, and for simple lag distributions easily analyse, such concepts as the mean lag, the lag "variance" etc. It has long been recognized[2] that the assumption of non-negativity is very arbitrary, and mainly introduced for computational convenience. We shall not go deeper into this, but just notice that the assumption of non-negativity facilitates the analysis somewhat, but is at the same time well known to be of limited realism.

The second assumption, that the weights sum to unity, is partly made to facilitate the view of the weights as probabilities, and partly for estimation purposes. If we try to estimate the expectations-augmented Phillips curve with $\pi_t^e$ formed adaptively,

$$\pi_t^e = f(u_t) + \alpha \sum_{j=1}^{n} w_j \pi_{t-j} + \varepsilon_t, \qquad (4.3)$$

---

[1] See for example Solow (1969) and Turnovsky (1972). Cf. also the survey in Gordon (1976, pp. 192-193).

[2] See e.g. Bailey (1962) and Griliches (1967).

it is only possible to estimate n of the (n+1) parameters $\alpha$, $w_1$, $w_2$, ..., $w_n$. This is so because the weighted sum of past inflation rates contains only n terms. Thus, in order to identify any of these (n+1) parameters, for example $\alpha$, or $w_3$, we must have some more information about them. We cannot really impose an exogenous value on $\alpha$, because this is just what we want to test; if $\alpha = 1$ the natural rate hypothesis holds for the economy (at least for the particular definition of equilibrium presented in Chapter 3 above) and whether this is true is the critical question to be answered by our estimation. The standard procedure is instead to assume that the adaptive weights sum to some constant value,

$$\Sigma w_j = k.$$

The particular value of k which is always used is unity[3] because with $\Sigma w_j = 1$ the expected inflation rate becomes an *average* of earlier inflation rates, and averages seem to have an intuitive appeal in statistical contexts. Another intuitively plausible reason for setting $\Sigma w_j = 1$ is that if inflation has always been constant, $\pi_{t-1} = \pi_{t-2} = \ldots = \pi$, then it is natural to set $\pi_t^e = \pi$, which conforms to the adaptive formula only if $\Sigma w_j = 1$.

Thomas Sargent has questioned the unity-sum constraint[4] by pointing out that *if* inflation has been at a constant rate for an infinite time it would admittedly be appropriate to set $\Sigma w_j = 1$, but in reality inflation *has not* been constant. Inflation has changed in a highly erratic way, and thereby

---

[3] For example, with a geometric (Koyck or Cagan) lag we have

$$\pi_t^e = (1-\delta) \sum_{j=0}^{\infty} \delta^j \pi_{t-1-j}.$$

From the formula for the sum of an infinite geometric series we immediately see that $\Sigma w_j = \Sigma(1-\delta)\delta^{j-1} = 1$.

[4] Sargent (1971).

the intuitive reason for constraining the weights is not applicable to the real world. Well then, if inflation has followed a stochastic process, shouldn't we set $\Sigma w_j = 1$ anyway? The answer is No, not in general. This is in fact one of the important cases where we obtain qualitatively different results depending on whether we choose to work with a stochastic or a deterministic model. The purpose of this chapter is to discuss and generalize Sargent's criticism of the weight sum constraint. We will show *firstly* that if we want to make an adaptive forecast of a *stationary stochastic process*, and if that forecast is to be optimal in the sense of minimizing the mean square error, then the adaptive weights should be *less than unity*. The conclusion will be that if people have believed that inflation has followed a stationary time series (and we saw in Chapter 2 above that there were some reasons for such a belief), the econometric studies referred to above might have obtained a downward bias in their estimates of the money illusion parameter $\alpha$. If the economic agents have regarded inflation as a stationary process, the imposed constraint that $\Sigma w_j = 1$ is inappropriate.

*Secondly*, we will discuss multi-period forecasting. The above expectations formula deals only with one-period-ahead forecasts, but a more general adaptive expectations mechanism is

$$\pi^e_{t+\theta} = \sum_{j=1}^{n} w_j \, \pi_{t-j}.$$

It has been claimed[5] that the optimal weight sum is a *decreasing* function of the forecasting horizon $\theta$. We will disprove this conjecture by a simple counterexample.

---

5 Sargent (1971).

## 4.1 ONE-PERIOD FORECASTING

Let us start with an illustrative example, a stationary stochastic process which for simplicity is assumed to be of the AR(1) type:

$$X_t = a\, X_{t-1} + \varepsilon_t \tag{4.4}$$

A typical realization of such a process, with $a = 0.9$, is shown in Figure 2.7 on page 25 above. Recalling the condition for stationarity of autoregressive processes, we know that (4.4) is stationary if and only if $|a| < 1$. Assume we want to make a forecast of the process on the adaptive form (4.4) with $n = 1$, i.e. with only one lagged observation:

$$X_t^e = wX_{t-1}.$$

Needless to say, the best forecast in terms of mean square error is given by

$$w = a.$$

It would thus be inappropriate to constrain the adaptive weight sum, which in this particular case consists of only one term, to sum to unity; such a constraint would only increase the mean square error.

This should make one quite suspicious of the unit-sum constraint, and Sargent (1971) in fact proved a slightly more general result: Assume we have a moving-average process

$$X_t = \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + \dots + \beta_p\varepsilon_{t-p} \tag{4.5}$$

(where the order p might well be infinite) such that $\beta_i \geq 0$ for all i. Suppose now that we want to write the process on the equivalent autoregressive form

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_q X_{t-q} + \varepsilon_t. \qquad (4.6)$$

Then, if the process is stationary it holds that

$$\sum_{j=1}^{q} \alpha_j < 1.$$

This result can fairly easily be proved.[6] Let us for simplicity introduce the lag, or back-shift, operator L, defined by $L^k X_t \equiv X_{t-k}$, and the lag polynomial[7]

$$B(L) = \beta_1 L + \beta_2 L^2 + \ldots + \beta_p L^p.$$

Our initial MA(p) process can then be written

$$X_t = [B(L) + 1]\varepsilon_t \qquad (4.7)$$

If we want to express (4.7) on the corresponding AR(q) form, we can write it similarly as

$$X_t = A(L) X_t + \varepsilon_t. \qquad (4.8)$$

We now want to characterize the lag polynomial $A(L) \equiv \alpha_1 L + \ldots + \alpha_q L^q$. Substitution of (4.7) into (4.8) yields

$$X_t = A(L)\ [B(L) + 1]\varepsilon_t + \varepsilon_t.$$

Thus,

$$B(L) = A(L)\ [B(L) + 1]$$

which, by setting L = 1, yields

---

6 Since Sargent's original proof is slightly more complicated I have chosen the following method of proof for expositional reasons.

7 For an account of the algebra of lag operators, see Dhrymes (1971, chapter 2).

$$A(1) = \frac{B(1)}{B(1) + 1}.$$

Since $A(1) \equiv \Sigma\alpha_j$ and $B(1) \equiv \Sigma\beta_j$ we thus have the desired result:[8]

$$\sum_{j=1}^{q} \alpha_j < 1.$$

The conclusion for adaptive forecasting is rather obvious: If we have a time series (4.5), or equivalently (4.6), of which we want to make a forecast

$$X_t^e = \sum_{j=1}^{n} w_j \; X_{t-k},$$

and if $n \geq q$, the best forecast is obtained by setting the adaptive weights equal to the autoregressive coefficients:

$$w_j = \alpha_j \;, \quad \text{all } j.$$

Since $\Sigma\alpha_j < 1$ if the time series is stationary, we also have that $\Sigma w_j < 1$. Now, Sargent pointed out that since all econometric studies had constrained the weights to sum to unity, while the economic agents might well have considered inflation as a non-explosive (i.e. stationary) process and formed optimal adaptive forecasts with $\Sigma w_j < 1$, there might have been a downward bias in the estimates of the money illusion parameter $\alpha$ in (4.3) which could explain the non-monetarist results obtained.

To judge whether this is plausible, we recapitulate Sargent's assumptions. We have

---

[8] Provided $\Sigma\beta_j \geq 0$, a requirement which is fulfilled by our initial assumption that $\beta_j \geq 0$, all j.

a) The time series must be of the type that can be expressed as a q-th order autoregressive process. This is an assumption we have to accept for the sake of mathematical simplicity. When discussing the Wold decomposition in Chapter 2 above, we noted that the processes that do not belong to this category are those with a *deterministic* sine or cosine wave involved; we choose to disregard these.

b) When expressed on the equivalent moving-average form (4.5), all MA-coefficients must be non-negative, or at least must their sum be non-negative. This is a limitation which might exclude important time series.

c) The time series must have a mean $E[X_t]$ equal to zero. This excludes most economic time series, at least in their raw form. The first or second differences might perhaps have a zero mean, but the economic models and the adaptive expectations formula have not in general dealt with such transformed series.

d) While the time series must be an AR(q) process according to point a) above, the adaptive formula (4.2) must be of an order n *at least as large as q*. But in reality, agents' memory (for which n is a proxy) is limited while perfectly reasonable time series might be of a very high order.

These assumptions are, taken together, quite restrictive. As an example of a time series to which Sargent's result is not applicable, we could take the Swedish inflation rates between 1861 and 1977. As shown in Chapter 2 above, they follow the ARMA process

$$\pi_t = 0.93 + 0.92\,\pi_{t-1} - 0.29\,\pi_{t-2} - 0.26\,\varepsilon_{t-4} + \varepsilon_t$$

or the equivalent MA($\infty$) process

$$\pi_t = 2.51 + \varepsilon_t + 0.92\,\varepsilon_{t-1} + 0.56\,\varepsilon_{t-2} + 0.25\,\varepsilon_{t-3} -$$

$$- 0.26\,\varepsilon_{t-4} - 0.28\,\varepsilon_{t-5} \ldots$$

This process is excluded by assumption b) above; when written on moving-average form, it contains both positive and negative coefficients. It is also excluded by assumption c) since its mean is 2.51, which is different from zero. Finally, if the adaptive expectations formula is of finite order, assumption d) is not satisfied; the process, written on autoregressive form, will be of infinite order.

We thus need a more general result concerning the optimal weight sum. In fact it turns out that the four assumptions above are not necessary for Sargent's conjecture to hold. We can state the following general theorem:

*Given an arbitrary time series* $\{\ldots, X_t, \ldots\}$ *of which we want to make a forecast*

$$X_t^e = \sum_{j=1}^{n} w_j X_{t-j}$$

*such that the mean square error is minimized. If the time series is stationary, then*

$$\sum_{j=1}^{n} w_j < 1.$$

The following three pages of this section contain a formal proof of that result[9]:

Consider the stationary process $\{\ldots, X_t, \ldots\}$, where $X_t$ is a real number. We define the second-moment function

$$r_{t-s} = r_{s-t} \equiv E[X_t X_s]$$

In a fashion similar to the way the covariance function is given a spectral representation, we can write the second-moment function as

---

9 The proof is rather cumbersome and can be skipped without loss of continuity. I am indebted to Anders Martin-Löf, who has worked out the mathematics for me.

$$r_\tau \equiv \int_{-\pi}^{\pi} e^{i\tau\lambda}\, dF(\lambda)$$

where $F(\lambda)$ is a mass distribution on $[-\pi,\pi]$. We want to make a forecast

$$X_t^e = \sum_{j=1}^{n} w_j\, X_{t-j} \equiv - \sum_{j=1}^{n} c_j\, X_{t-j},$$

where $c_j \equiv -w_j$ and where the $c_j$ weights are such that $V \equiv E\,(X_t - X_t^e)^2$ is minimized. By setting $c_0 \equiv 1$, we have

$$V \equiv E[(X_t - X_t^e)^2] = E\left[\left(\sum_{j=0}^{n} c_j X_{t-j}\right)^2\right] =$$

$$= \sum_{j=0}^{n}\sum_{k=0}^{n} c_j c_k E[X_{t-j}X_{t-k}] = \sum_{j=0}^{n}\sum_{k=0}^{n} c_j c_k r_{j-k} =$$

$$= \int_{-\pi}^{\pi}\sum_{j=0}^{n}\sum_{k=0}^{n} c_j c_k e^{i\lambda(j-k)} dF(\lambda) =$$

$$= \int_{-\pi}^{\pi}\left|\sum_{j=0}^{n} c_j e^{i\lambda(n-j)}\right|^2 dF(\lambda).$$

Thus, $(c_1, c_2, \ldots, c_n)$ should be chosen so as to minimize

$$\int_{-\pi}^{\pi}\left|e^{in\lambda} + c_1 e^{i(n-1)\lambda} + \ldots + c_{n-1}e^{i\lambda} + c_n\right|^2 dF(\lambda).$$

We introduce the polynomial

$$c(z) \equiv z^n + c_1 z^{n-1} + \ldots + c_n$$

and can then write this expression as

$$\int_{-\pi}^{\pi} c(e^{i\lambda})\, dF(\lambda).$$

Now, the idea is to express the optimal predictor in terms of a particular orthonormal set of polynomials. Then we shall use a property of these polynomials to show our main result concerning the weight sum. Let us regard the space of polynomials

$$P(z) = \sum_{t=0}^{n} p_t z^t \qquad Q(z) = \sum_{s=0}^{n} q_s z^s$$

with the inner product

$$(P,Q) \equiv \int_{-\pi}^{\pi} P(e^{i\lambda}) \overline{Q(e^{i\lambda})} \, dF(\lambda) \equiv$$

$$\equiv \int_{-\pi}^{\pi} \sum_t \sum_s p_t \overline{q_s} \, e^{it\lambda} e^{-is\lambda} \, dF(\lambda)$$

for some given function $F(\lambda)$. If we take F to be the spectral distribution of our stochastic process $\{X_t\}$, we thus have

$$(P,Q) = \sum_t \sum_s p_t \overline{q_s} \, r_{t-s}$$

Choose a particular sequence of vectors in this space, namely the sequence of powers of z

$$1,\ z,\ z^2,\ z^3,\ \ldots$$

We then have, according to our definition of inner product,

$$(z^t, z^s) = \int_{-\pi}^{\pi} e^{it\lambda} \overline{e^{is\lambda}} \, dF(\lambda) = r_{t-s} = r_{s-t}.$$

The sequence of vectors $\{z^t\}$ can be orthogonalized by the Gram-Schmidt Procedure[10], i.e., we can obtain an orthonormal sequence of polynomials $\{\phi_t(z)\}$ where

---

[10] Cf. Luenberger (1969), Section 3.5.

$$\phi_t(z) = \sum_{k=0}^{t} \phi_{t,k} z^k \qquad \phi_{t,t} > 0$$

with the inner product

$$(\phi_t, \phi_s) = \int_{-\pi}^{\pi} \phi_t(e^{i\lambda}) \overline{\phi_s(e^{i\lambda})} \, dF(\lambda) = \delta_{t,s}$$

For our proof, we now need a particular result which we state as

*Lemma:* The optimal prediction is given by

$$c(z) \equiv z^n + c_1 z^{n-1} + \ldots + c_n = \frac{\phi_n(z)}{\phi_{n,n}} .$$

*Proof:* An arbitrary n-th degree polynomial $c(z) = z^n + c_1 z^{n-1} + \ldots + c_n$ can be written on the equivalent form

$$c(z) = a_o \phi_n(z) + a_1 \phi_{n-1}(z) + \ldots + a_n \phi_o(z)$$

with $a_o = \frac{1}{\phi_{n,n}}$ . We then have

$$\int_{-\pi}^{\pi} \left(c(e^{i\lambda})\right)^2 dF(\lambda) = \left(c(z), c(z)\right) =$$

$$= \sum_{j=0}^{n} \sum_{k=0}^{n} a_j a_k (\phi_{n-j}, \phi_{n-k}) = \sum_{j=0}^{n} a_j^2$$

since the polynomials $\phi_n$ are orthonormal. $a_o$ is given by $a_o = \frac{1}{\phi_{n,n}}$ , but all other $a_j$ coefficients are free at our disposal. Thus,

$$V \equiv \int \left(c(e^{i\lambda})\right)^2$$

is minimized if we set

$$a_1 = a_2 = \ldots = a_n = 0$$

and consequently

$$c(z)\ c(z) = a_o\phi_n(z) = \frac{\phi_n(z)}{\phi_{n,n}}$$

This completes the proof of the Lemma.

We can now use this result to show that

$$w_1 + w_2 + \ldots + w_n \equiv -(c_1 + c_2 + \ldots + c_n) < 1.$$

We have $c(1) \equiv 1 + c_1 + c_2 + \ldots + c_n$. That is, if $c(1) > 0$, then $\Sigma w_j < 1$. By the Lemma,

$$c(z) = \frac{\phi_n(z)}{\phi_{n,n}} \qquad \text{and } \phi_{n,n} > 0.$$

The question is thus whether $\phi_n(1) > 0$. And it is a known result[11] that $\phi_n(z)$ has all its zeros within the unit circle in the complex plane, i.e., $|\zeta_i| < 1$, all i. If we let z increase from 1 to infinity, $\phi_n(z)$ will thus never pass zero, and for large z values $z^n$ will dominate, that is $\phi_n(z) \approx \phi_{n,n}z^n > 0$ for large z. Consequently, $\phi_n(1) > 0$, which means that $c(1) > 0$. This completes the proof.

Now, with this result in mind we can see that Sargent's conjecture was justified: if people form adaptive expectations (as they have been assumed to do in the econometric studies referred to above) and if they are rational in the particular sense of trying to make as good adaptive forecasts as possible, they will use $w_j$ weights summing to less than unity, provided that they regard the time series to be forecasted *as stationary*. An econometric study which incorrectly constrains the weights to sum to unity would therefore yield biased estimates of other parameters, for example the money illusion $\alpha$ in the expectations-augmented Phillips curve.

---

[11] See Akhiezer (1965), p. 184.

Also, this result sheds some light on why Muth (1960), in his work on optimal adaptive forecasting, obtained predictors of the form

$$Y_t^e = \delta \sum_{j=1}^{\infty} (1-\alpha)^{j-1} X_{t-j}$$

where the weight sum is obviously equal to unity. An examination of Muth's results shows that he did not study stationary processes; the time series represented by equations (3.21) and (3.23) in the previous chapter are non-stationary, modified random walks. The same holds for Nerlove's (1967) paper, where the unobservable variable $X_t$ is forecasted according to

$$X_t^e = \frac{\psi - \beta}{\psi} \sum_{j=0}^{\infty} \beta^j Y_{t-j}.$$

If the underlying process $\{X_t\}$ is stationary, i.e. $\psi < 1$,[12] we see that the weight sum will be less than unity. If we let $\psi$ approach unity, i.e. we allow the underlying process to become more and more like a random walk, the optimal weight sum will approach unity too.

## 4.2 MULTI-PERIOD FORECASTING

The above analysis is confined to one-period-ahead forecasts, a time span which is almost always assumed in macroeconomic expectations models. In the attempts to provide micro-foundations for the Phillips curve, the models have in general dealt only with optimization over two adjacent periods. In reality, however, individuals often enter multi-period contracts.[13] If an economy with such contracts is to be represented by a Phillips curve, we have perhaps the relation

---

[12] Cf. page 59 above.

[13] For a macro model with multi-period contracts, which however deals with other questions than the ones discussed in this paper, see Fischer (1977).

$$\pi_t = f(u_t) + \alpha \cdot \pi^e_{t+\theta} + \varepsilon_t \tag{4.9}$$

In general it is of course not the expected inflation rate at one single future date that matters, but at several dates:

$$\pi_t = f(u_t) + \phi(\pi^e_t, \pi^e_{t+1}, \pi^e_{t+2}, \ldots) + \varepsilon_t.$$

For simplicity we disregard this complication, and concentrate on an economy represented by (4.9), where all future periods have been collapsed into the single period $t+\theta$. Assume we would like to estimate (4.9) with adaptive expectations. We then have

$$\pi_t = f(u_t) + \alpha \sum_{j=1}^{n} w_j \pi_{t-j} + \varepsilon_t \tag{4.10}$$

which looks exactly similar to its one-period-ahead counterpart (4.3). Thus, when estimating a Phillips curve model, it seems like we do not need to know whether it represents a micro behavior with one-period contracts or with multi-period contracts, i.e. if $\Sigma w_j \pi_{t-j}$ should be a proxy for $\pi^e_t$ or $\pi^e_{t+\theta}$ with $\theta > 0$. However, if the forecasting horizon implies different constraints on the optimal weight sum, the forecasting horizon matters. In this section we will briefly discuss two conjectures, due to Sargent (1971):

i) The optimal weight sum is less than unity for multi-period forecasting.

ii) The optimal weight sum is a decreasing function of the forecasting horizon.

Multi-period forecasting is substantially more complicated than the one-period forecasting studied in the previous section, and we can thus not apply as general methods. However, by simple, numerical counterexamples we will show that both these conjectures are wrong.

Assume we have a stationary time series $\{X_t\}$ of which we want to make an optimal prediction of the form

$$X^e_{t+\theta} = \sum_{j=1}^{n} w_j \; X_{t-j}.$$

Optimal linear predictions of stationary stochastic processes can be looked upon in two equivalent ways. Either we regard the optimization as a problem of minimizing a suitable norm in a linear space, i.e. by finding a vector of adoptive $w_j$ weights which yields residuals orthogonal to the time series.[14] Or we choose the equivalent, but less abstract, way of finding the adaptive weights by simply taking the derivative of the mean square error,

$$V \equiv E[(X_{t+\theta} - \Sigma w_j X_{t-j})^2],$$

with respect to $w_j$, $j=1,\ldots,n$, equal to zero.[15] In both cases we obtain the vector of optimal weights by solving the equation system

$$\begin{pmatrix} r_0 & r_1 & r_2 & \cdot & \cdot & \cdot & r_{n-1} \\ r_1 & r_0 & r_1 & \cdot & \cdot & \cdot & r_{n-2} \\ r_2 & r_1 & r_0 & \cdot & \cdot & \cdot & r_{n-3} \\ \cdot & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot \\ r_{n-1} & r_{n-2} & r_{n-3} & \cdot & \cdot & \cdot & r_0 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ \cdot \\ \cdot \\ \cdot \\ w_n \end{pmatrix} = \begin{pmatrix} r_{\theta+1} \\ r_{\theta+2} \\ r_{\theta+3} \\ \cdot \\ \cdot \\ \cdot \\ r_{\theta+n} \end{pmatrix}$$

where $r_\tau$ is the second moment, defined by

$$r_\tau \equiv E[X_t \cdot X_{t+\tau}].$$

This system can be more compactly written

$$Rw = r(\theta). \tag{4.11}$$

---

[14] Cf. Luenberger (1969).

[15] See e.g. Cox and Miller (1965, chapter 7) or Karlin and Taylor (1975, chapter 9).

This equation system, which gives the optimal adaptive weights in terms of the second moments, will in the following be referred to as the *orthogonality conditions*. Recalling the Yule-Walker equations $\Gamma\alpha = \gamma$ on page 24 in Chapter 2, we see that they look quite similar in structure. In fact, if $E[X_t] = 0$ the second-moment matrix R is identical to the covariance matrix $\Gamma$, and the orthogonality conditions (4.11) for one-step-ahead forecasts (i.e. $\theta = 0$) become identical to the Yule-Walker equations.

Now, the advantage of analyzing the optimal w vector by (4.11) instead of obtaining it by the frequency domain approach of the previous section is that equation (4.11) is so simple and clear-cut. Numerical computations should therefore be carried out by means of the orthogonality conditions. Analytical results, however, are perhaps more difficult to reach this way. While we know from Section 4.1 that, for one-period forecasts, it must hold that

$$a^T w = a^T r(0) R^{-1} < 1$$

(where $a^T \equiv (1, 1, \ldots, 1)$), this is not easily shown by analyzing the r(0) vector and the R matrix alone.[16] And for multi-period forecasts the conjecture that

$$e^T R^{-1} r(\theta) < 1$$

---

16 For some simple cases, however, it can be done. Assume $\theta = 0$ and the order of the lag, n, is equal to 2. Then R will be a 2x2 matrix, and solving (4.11) for w yields, after some computations,

$$w_1 + w_2 = \frac{r_1 + r_2}{r_1 + r_0}$$

Since the second-moment function has the property that $|r_\tau| < r_0$ for all $\tau \neq 0$, this expression is obviously less than unity.

are difficult to prove (or disprove) analytically for non-trivial cases.[17] However, since the strength in the orthogonality approach lies in the computational simplicity provided by (4.11), we can try to disprove the conjecture by searching for counter-examples. For n = 2, for example, a counter-example should satisfy

$$\frac{r_0 r_{\theta+1} - r_1 r_{\theta+2} - r_1 r_{\theta+1} + r_0 r_{\theta+2}}{r_0^2 - r_1^2} \geq 1 \qquad (4.12)$$

Such a process can easily be constructed. Assume that we make two-period forecasts, i.e. $\theta = 1$. Then (4.12) is satisfied by any process with, e.g.,

$r_0 = 2.5$

$r_1 = -2$

$r_2 = 1$

$r_3 = 0$

$r_\tau$ = arbitrary for $\tau > 3$.

For these figures we would have $\Sigma w_j = 2$. A process with these second moments can easily be found, e.g. the AR(4) process

$$X_t = -1.33\ X_{t-1} - 0.44\ X_{t-2} + 0.44\ X_{t-3} + 0.33\ X_{t-4} + \varepsilon_t.$$

$$\sigma_\varepsilon^2 = 2.06$$

as can be checked by applying the Yule-Walker equations.[18] Solving the characteristic equation shows that this process is actually stationary, as was to be required. In fact there is an infinite number of ARMA processes satisfying (4.12).

---

[17] If n = 1 (4.11) gives the optimal w (now a scalar) by $w = r_{\theta+1}/r_0$, which is obviously always less than unity.

[18] In this case $r_4$ is set equal to -0.5.

Recalling the two conjectures mentioned on page 88 we see that the first one has been disproved; the sum of the weights is not necessarily less than unity for multi-period forecasting. The second conjecture, which has been claimed true by Sargent (1971, page 725), stated that the weight sum was a decreasing function of the forecasting horizon. This has actually already been disproved by the above example, since $\Sigma w_j$ for one-period forecasts is always less than unity according to the general result in Section 4.1 above, and equal to two for two-period forecasts according to the numerical example. Thus it cannot be a decreasing function of the forecasting horizon.

Another illustrative example is provided by the simple case with an adaptive formula of order n = 1. Then the optimal weight is given by

$$w = \frac{r_{\theta+1}}{r_0} .$$

We see thus that although w is strictly less than unity for all $\theta$, it will be a monotonically decreasing function of $\theta$ *only if* $r_{\theta+1}$ is monotone, which is far from the general case. Thus, neither of the two conjectures stands to closer examination, and the conclusion is that multi-period forecasting is so cumbersome that simple and clear-cut results are hard to obtain.

## 4.3 CONCLUDING COMMENTS

In this chapter we have demonstrated that if one wants to make an adaptive one-period forecast which is optimal in the sense of minimizing the mean square error, and if the variable to be forecasted follows a stationary process, then the sum of the adaptive weights should be less than unity. This result seems slightly counterintuitive, and one is tempted to ask whether such a forecasting formula doesn't mean that on the average, the forecasts will underestimate the variable to be forecasted. The answer is that this might happen; in other words, there might be a bias in the forecasts. For a predictor $X_t^e$ to be unbiased, we require that

$$E[X_t^e] = E[X_t].$$

If $X_t^e$ is formed adaptively we thus have for a stationary series

$$E[\Sigma w_j X_{t-j}] \equiv \Sigma w_j E[X_t] = E[X_t],$$

which holds only if $\Sigma w_j = 1$ or $E[X_t] = 0$. Thus, for time series with $E[X_t] \neq 0$, optimal adaptive forecasts will be biased. When having to choose between optimal but biased forecasts, and inoptimal (i.e. with $\Sigma w_j = 1$) but unbiased, the choice is not very straightforward. The dilemma stems from two problems regarding the concept of optimal adaptive expectations.

*Firstly* we have the choice of criterion for optimality. We have assumed that "optimal" means "minimizing the mean square error", and this is the dominating - in fact the only - criterion of optimality used in the forecasting and econometric literature. If we instead of minimizing the expected value of the *square* of the residuals had chosen to minimize e.g. the expected value of a *linear* function of the residuals

$$V' \equiv E[X_t - X_t^e],$$

the problem of bias would never have occurred. There are however several arguments against such an optimality criterion, as can be seen in the introductory chapter of most econometric textbooks, and we have chosen to discuss only the criterion dealt with in the literature.

*Secondly*, we have the question of whether it is reasonable to make adaptive forecasts at all. This can of course be discussed, but as long as such expectations are used in economic and econometric studies, their properties should clearly be investigated. A way to escape from the bias in the forecast would be to add a constant term in the formula, and make forecasts according to

$$x_t^e \equiv k + \Sigma w_j \, x_{t-j}.$$

By an appropriate choice of k, such a formula can always be guaranteed to yield unbiased forecasts. With this formula, however, we have left the concept of adaptive expectations (which was the topic we had chosen for the present chapter) behind us, and are approaching the concept of *rational* expectations (which will be the topic of Chapters 6 and 7 below).

# 5. THE NATURAL RATE HYPOTHESIS IN A MODEL WITH ADAPTIVE EXPECTATIONS

The result of the previous chapter - that the adaptive weights should have a sum which is less than unity - seems at first sight to have most relevance for empirical estimation. However, it has equally important implications for the theoretical properties of macro models, and some of these will be the topic of this chapter. In particular, we are used to the result that if there is no money illusion and if expectations are formed adaptively in the standard way (i.e. with $\Sigma w_j = 1$), then the long-run Phillips curve is vertical. This is so because if inflation is given by the relation

$$\pi_t = f(u_t) + \Sigma w_j \pi_{t-j} + \varepsilon_t$$

and if $\{\pi_t\}$ is stationary, we can take the expectation of both members to obtain

$$(1-\Sigma w_j)E[\pi_t] = E[f(u_t)]$$

which means that if $\Sigma w_j = 1$, then $E[f(u_t)]$ must be zero for finite values of $E[\pi_t]$. This can intuitively be interpreted as a vertical long-run Phillips curve. Similarly, we can reformulate the relation as

$$E[\pi_t] = \frac{E[f(u_t)]}{1 - \Sigma w_j}$$

which shows that if $\Sigma w_j = 1$ and $E[f(u_t)]$ is different from zero then $E[\pi_t]$ is undefined. However, if $\Sigma w_j$ is less than unity, then there might be some trade-off between $E[\pi_t]$ and $E[f(u_t)]$, which can be interpreted as a long-run Phillips trade-off. Now all this has been highly intuitive, and for a satisfactory analysis of whether the long run Phillips curve is vertical or not when the adaptive expectations are formed optimally, we have to make some formal definitions.

We will deal with the two standard approaches to represent the economy that have been presented in Chapter 3 above, namely the Phillips curve and the aggregate supply equation. As has already been mentioned, they need not be mutually exclusive; it is possible to obtain the former from models building on the latter. On the other hand, they *could* be regarded as fundamentally different concepts - the former often relating to a Keynesian disequilibrium world, while the latter is of a neoclassical equilibrium breed. However, they turn out to have the same formal structure, and *for the present purpose* they could thus be analyzed interchangeably. To clarify this, we repeat the Phillips curve

$$\pi_t = f(u_t) + \alpha\, \pi_t^e + \varepsilon_t,$$

set the money illusion parameter $\alpha = 1$, and define the disturbance $v_t \equiv - \varepsilon_t$ to obtain

$$f(u_t) = (\pi_t - \pi_t^e) + v_t.$$

This looks very similar to the supply equation

$$Y_t = a + b(P_t - P_t^e) + \eta_t$$

and from the formal point of view they could be analyzed by the same methods. For expositional reasons I will concentrate on the Phillips curve formulation and disregard the supply function, but those who do believe in the latter rather than in the former can equally well look upon the following pages as if they dealt with the supply function instead.

## 5.1 THE FORMULATION OF A MACRO MODEL, AND THE DEFINITION OF A "NATURAL RATE".

We assume that the economy can be represented by the Phillips curve, and we assume that there is no money illusion, i.e. that $\alpha = 1$:

$$\pi_t = f(u_t) + \pi_t^e + \varepsilon_t \tag{5.1}$$

The authorities are assumed to control the unemployment rate via public works and different employment policies; an *economic policy* is thus defined as a time series $\{\ldots, u_{t-1}, u_t, u_{t+1}, , \ldots\}$. For a given expectations mechanism

$$\pi_t^e = \sum_{j=1}^{n} w_j \, \pi_{t-j} \tag{5.2}$$

this policy is transmitted through the economic system with an inflationary process $\{\ldots, \pi_{t-1}, \pi_t, \pi_{t+1}, \ldots\}$ as the result.

In our discussion at the end of Chapter 2 of the equilibrium concept, we said that there are three aspects of equilibrium: equality of supply and demand, consistency, and stationarity. The first aspect will not be dealt with in this macro context; since the micro model underlying the Phillips curve might well be a Keynesian disequilibrium model, equality between supply and demand is hardly relevant to the problems of the present chapter. The second aspect, however, is essential. With consistency we mean that the forecasts $\pi_t^e$ formed by (5.2) are *optimal* forecasts (in terms of mean square error) of the inflationary process $\{\pi_t\}$ resulting from our model. By stationary we mean, quite naturally, that the inflationary process resulting from this (consistent) model is stationary. Thus, *a dynamic equilibrium* is a stationary time series $\{u_t\}$, a stationary time series $\{\pi_t\}$, which is related to $\{u_t\}$ according to the model (5.1), and an expectations mechanism (5.2) which yields optimal expectations of $\{\pi_t\}$.

If $\{\pi_t\}$ and $\{u_t\}$ form stationary time series, their averages $E[\pi_t]$ and $E[u_t]$ are constant. Plotted in a diagram, these averages can be interpreted in terms of the long-run Phillips curve:

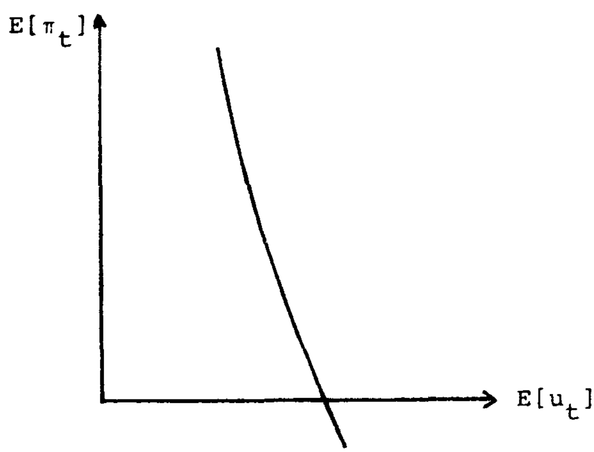


Figure 5.1: The long-run Phillips curve.

This diagram depicts the set of dynamic equilibria; it tells us which inflation rate will result *on the average* for a given average of the unemployment rate. In the figure it has been depicted as downward-sloping, but it might equally well be vertical. In such a case the natural rate hypothesis holds: $E[u_t]$ cannot deviate from a fixed, "natural" level $u^N$ without making the average inflation rate undefined (or infinite).

We can thus state our formal definition of the natural rate hypothesis (NRH) as follows: If there exists a constant number $u^N$ such that $E[u_t]$ is equal to $u^N$ for all stationary time series $\{\pi_t\}$, and if $E[u_t] \neq u^N$ only for undefined $E[\pi_t]$, then the NRH is said to hold for this model.

There is, however, a small problem with this definition, which has to do with the *curvature* of the function $f(u_t)$. It is well known from e.g. Lipsey's (1960) paper that the authorities can exploit the curvature by changing the variance in the $\{u_t\}$ process, thereby affecting $E[\pi_t]$.[1] Now this complication is inessential to our main point, and we will simply disregard it by making a slight reformulation of our definition of economic policy. Instead of defining a policy as a (stationary) time series $\{\ldots, u_{t-1}, u_t, u_{t+1}, \ldots\}$ we define a policy as a (stationary) time series $\{\ldots, f(u_{t-1}), f(u_t), f(u_{t+1}), \ldots\}$ or more briefly $\{\ldots, f_{t-1}, f_t, f_{t+1}, \ldots\}$. Thus our definition of the natural rate should be reformulated:

*If there exists a constant number $f^N$ such that $E[f_t]$ is equal to $f^N$ for all stationary time series $\{\pi_t\}$, and if $E[f_t] \neq f^N$ only for undefined $E[\pi_t]$, then the NRH is said to hold for this model.*

Note that we have formulated the NRH in terms of whether $E[\pi_t]$ is defined or not; we have thus not mentioned the concept of stationarity in the $\{\pi_t\}$ process. However, these two ways of regarding the NRH are equivalent. We will deal only with inflationary processes that can be written on ARMA form. For such processes, stationarity means that $E[\pi_t]$ is finite, and non-stationarity means that the process explodes (or drifts away like the random walk), thereby making the unconditional mean $E[\pi_t]$ undefined. Thus the NRH can be analyzed *eihter* in terms of the mean of the inflation process, *or* in terms of the stationarity of the process; the two concepts are equivalent.

---

[1] Lipsey's point was related to the variance over markets, and is thus conceptually different from our variance *over time*, but the same principle naturally carries over to our time series context.

## 5.2 THE LONG-RUN IMPACT OF ECONOMIC POLICY

In this section we will approach the question of whether there exists a stationary time series $\{\pi_t\}$ for which $E[f_t]$ can be different from its "natural" value $f^N$ (which we in fact know is equal to zero). For the sake of simplicity we make two assumptions, the first of which can easily be dispensed with, and the second being of more crucial importance.

*Firstly*, we will assume, in the original Phillips curve tradition, that the direction of causality runs *from* $\{u_t\}$, via the relation

$$\pi_t = f(u_t) + \sum_{j=1}^{n} w_j \pi_{t-j} + \varepsilon_t$$

*to* $\{\pi_t\}$. The authorities have full control (through public works etc.) over the $u_t$ time series; needless to say, they thereby also control the $f_t$ process. Now we postulate for simplicity that the authorities restrict themselves to deterministic $f_t$ processes, i.e. they peg $u_t$ at a certain value u, such that $f_t$ is also pegged to a constant f, not necessarily equal to the natural value $f^N = 0$. And the question is: Suppose $f \neq 0$. Will then the autoregressive process

$$\pi_t = f + \sum_{j=1}^{n} w_j \pi_{t-j} + \varepsilon_t \tag{5.3}$$

where the $w_j$ weights are such that they give an optimal adaptive forecast of the process itself, be stationary or not?

The assumption of constancy in the $f_t$ series is perhaps not quite realistic; even if the authorities can control $\{u_t\}$, they can perhaps not control it completely. The assumption facilitates the analysis considerably, however, and we thus have to accept it. However, it can be slightly modified if preferred: The authorities do not control $f_t$ completely, but have to recognize a disturbance which forms a white noise process. Thus,

$$f_t = f + \eta_t. \tag{5.4}$$

Inserting this into the Phillips curve, we have

$$\pi_t = f + \sum_{j=1}^{n} w_j \pi_{t-j} + e_t \tag{5.3'}$$

where the disturbance $e_t \equiv (\varepsilon_t + \eta_t)$ is white noise. This is the same kind of function as (5.3); with this in mind, (5.3) might seem more realistic.

*The second assumption* is perhaps more difficult to accept. It says that the economic agents immediately "know" the true structure of the stochastic process they are forecasting. Or, alternatively, that the agents make estimates before they know the structure of the process - but we disregard the transition period during which the estimates are made and confine our analysis to the "equilibrium" time when the process is known. This is of course a limitation of our analysis, but it is necessary in order to make the mathematics manageable. It is also in conformity with the practice in the rational expectations literature which disregards the transition period.[2]

The specific nature of this assumption can be more illuminated if we state clearly how the dynamic process of an active macro policy works. For notational simplicity we introduce the lag polynomial

$$W \equiv w_1 L + w_2 L^2 + \ldots + w_n L^n,$$

---

[2] For exceptions, see Taylor (1976) and B. Friedman (1979).

where L is the shift operator defined by $L^k X_t \equiv X_{t-k}$. Equation (5.3) can thus be written

$$\pi_t = f + W \pi_t + \varepsilon_t. \tag{5.5}$$

Now we assume that the economy is initially in a state of dynamic equilibrium with unemployment at its natural rate, so that f = 0, and inflation forming a white noise process

$$\pi_t = \varepsilon_t. \tag{5.6}$$

This initial state is a dynamic equilibrium (cf. the definition on page 98 above) in the threefold sense that inflation is stationary, that it conforms to (5.5) (with $w_i = 0$ for all i), and that such a w vector actually yields an optimal adaptive forecast of the process itself. Let us denote this initial process by $\{\pi_t\}^0$, and the lag polynomial associated with the initial $w_i$ vector by $W^0$. The process $\{\pi_t\}^0$ can thus be written either on the form (5.6) or on the equivalent form

$$\pi_t = W^0 \pi_t + \varepsilon_t. \tag{5.6'}$$

Now the authorities decide to peg unemployment at a rate different from the natural one, so that f becomes pegged to a value other than zero. This means that $\{\pi_t\}$ is not given by (5.6') any longer, but by (5.6') *plus* a constant $f \neq 0$:

$$\pi_t = f + W^0 \pi_t + \varepsilon_t. \tag{5.7}$$

Let us denote the process by $\{\pi_t\}^1$. This process, however, does not necessarily constitute a dynamic equilibrium, since the polynomial $W^0$ does not necessarily yield optimal adaptive predictions of (5.7). The economic agents soon discover that the inflationary process has changed from (5.6') to (5.7) and start to make optimal adaptive forecasts of the latter:

$$\pi_t^e = W^1 \pi_t.$$

The lag polynomial $W^1$ is thus the polynomial which gives an optimal adaptive prediction of the process $\{\pi_t\}^1$. Since the agents have revised their inflationary expectations, inflation will change accordingly and follow the new process $\{\pi_t\}^2$ given by

$$\pi_t = f + W^1 \pi_t + \varepsilon_t.$$

The agents will then revise their forecasts and form

$$\pi_t^e = W^2 \pi_t,$$

which in turn gives rise to the process $\{\pi_t\}^3$,

$$\pi_t = f + W^2 \pi_t + \varepsilon_t,$$

which is optimally forecasted by

$$\pi_t^e = W^3 \pi_t$$

and so on. The question is now: does the sequence of polynomials $\{\ldots, W^{\nu-1}, W^{\nu}, W^{\nu+1}, \ldots\}$, and the corresponding sequence of processes

$$\left\{\ldots, \{\pi_t\}^{\nu}, \{\pi_t\}^{\nu+1}, \{\pi_t\}^{\nu+2}, \ldots\right\},$$

converge to some polynomial $W^*$, and some equilibrium process $\{\pi_t\}^*$? The process $\{\pi_t\}^*$, more explicitly written as

$$\pi_t = f + W^* \pi_t + \varepsilon_t,$$

is an equilibrium if the polynomial $W^*$ yields optimal predictions of the process (and this is always the case if the sequence $\{W^{\nu}\}$ has converged), and if it is stationary. If there exists an equilibrium process $\{\pi_t\}^*$ for $f \neq 0$, then the NRH does not hold for our model. Conversely, if no stationary $\{\pi_t\}^*$ process exists unless $f = 0$, the NRH holds.

The next section will deal with these questions, and it will then be shown that all processes $\{\pi_t\}^\nu$ are actually stationary for finite $\nu$. However, if we let $\nu \to \infty$, it will turn out that although a process $\{\pi_t\}^*$ exists, it is not stationary. The sequence

$$\left\{\ldots, \{\pi_t\}^{\nu-1}, \{\pi_t\}^{\nu}, \{\pi_t\}^{\nu+1}, \ldots\right\}$$

converges to a *non-stationary, random-walk-like process with drift* if $f \neq 0$. The economic interpretation of this is that the NRH, as defined above, holds for our model.

## 5.3 CONVERGENCE OF INFLATIONARY PROCESSES

Let us introduce a few notations. We have the lag polynomial

$$W^\nu \equiv w_1^\nu L + w_2^\nu L^2 + \ldots + w_n^\nu L^n.$$

Let us denote the vector of coefficients associated with this polynomial by $w^\nu$,

$$w^\nu \equiv (w_1^\nu \; w_2^\nu \; \ldots \; w_n^\nu)^T,$$

and the sum of the coefficients by $\Sigma^\nu$:

$$\Sigma^\nu \equiv \sum_{j=1}^{n} w_j^\nu.$$

In this section we will show that the sequence of vectors $\{\ldots w^{\nu-1}, w^\nu, w^{\nu+1}, \ldots\}$ converges to a vector $w^*$. Further, we will show that the sequence of coefficient sums $\{\ldots, \Sigma^{\nu-1}, \Sigma^\nu, \Sigma^{\nu+1}, \ldots\}$ converges to a scalar $\Sigma^* = 1$. Finally, we will discuss what this means in terms of convergence of the inflationary processes

$$\pi_t = f + W^{\nu} \pi_t + \varepsilon_t .$$

To demonstrate the convergence of the vectors $w^{\nu}$ we proceed by means of a few Lemmas:

*Lemma 1:* Given a stationary AR(n) process

$$\pi_t = f + \sum_{j=1}^{n} \alpha_j \pi_{t-j} + \varepsilon_t$$

of which we want to make an optimal adaptive forecast

$$\pi_t^e = \sum_{j=1}^{n} w_j \pi_{t-j}$$

It then holds that $\Sigma w_j \geq \Sigma \alpha_j$, with strict inequality if $f \neq 0$. Further, if $\alpha_j \geq 0$, all j, then $w_j \geq \alpha_j$, all j.

*Lemma 2:* Given an AR(n) process

$$\pi_t = f + \sum_{j=1}^{n} w_j \pi_{t-j} + \varepsilon_t .$$

If $w_j \geq 0$, all j, and $\Sigma w_j < 1$, then the process is stationary.

*Lemma 3:* Given an AR(n) process

$$\pi_t = f + \sum_{j=1}^{n} w_j \pi_{t-j} + \varepsilon_t .$$

If $w_j \geq 0$, all j, and $\Sigma w_j = 1$, then the process is non-stationary; the roots $\zeta_i$ of its characteristic equation lie on and outside the unit circle in the complex plane, i.e.

$$|\zeta_i| \geq 1 \quad \text{with equality for at least one i.}$$

The string of Lemmas[3] leads us to the desired result, namely that the sequence of vectors $w^{\nu}$ converges to a vector $w^*$, and that the process associated with $w^*$ is what we have re-referred to above as "random-walk-like".

---

[3] The proofs will be given at the end of this section.

We start with the initial, equilibrium process $\{\pi_t\}^0$, defined by (5.6'). This process is obviously stationary, and so is the next process, $\{\pi_t\}^1$, defined by (5.7). The coefficient vector $w^0$ of the process $\{\pi_t\}^1$ is obviously non-negative; the process therefore satisfies the assumptions of Lemma 1. We thus know that it can be optimally predicted by

$$\pi_t^e = W^1 \pi_t$$

where $W^1$ has the property that $\Sigma^0 < \Sigma^1$ and that $0 \leq w^0 \leq w^1$. Further, we know from the result on the optimal weight sum in the previous chapter that $\Sigma^1 < 1$.

The next process is $\{\pi_t\}^2$, defined by

$$\pi_t = f + W^1 \pi_t + \varepsilon_t.$$

We know from Lemma 1 that all coefficients are non-negative, and that their sum is less than unity. Thus $\{\pi_t\}^2$ satisfies the assumptions of Lemma 2 and is therefore stationary. But if $\{\pi_t\}^2$ is stationary we can apply Lemma 1; the process can hence be predicted by

$$\pi_t^e = W^2 \pi_t$$

where the polynomial $W^2$ is such that $\Sigma^0 < \Sigma^1 < \Sigma^2$ and that $0 \leq w^0 \leq w^1 \leq w^2$. Further, from our earlier result on the optimal weight sum we also have that $\Sigma^2 < 1$. For the next process, $\{\pi_t\}^3$, the same reasoning holds; we have for the general process $\{\pi\}^\nu$:

$$\Sigma^0 < \Sigma^1 < \ldots < \Sigma^\nu < \Sigma^{\nu+1} < \ldots < 1$$

and

$$0 \leq w^0 \leq w^1 \leq \ldots \leq w^\nu \leq w^{\nu+1} \leq \ldots$$

The sequence of scalars $\{\ldots, \Sigma^{\nu}, \ldots\}$ is thus monotonically increasing and bounded above. Hence it converges to a scalar $\Sigma^*$. Similarly, the sequence of non-negative vectors $\{\ldots, w^{\nu}, \ldots\}$ is monotonically increasing. It is also bounded above since the sum of the elements $\Sigma w_j^{\nu} \equiv \Sigma^{\nu}$ is bounded above by $\Sigma^*$. Thus this sequence converges in $R^n$ to a limit vector $w^*$.

We will prove by contradiction that $\Sigma^* = 1$. Assume this is not the case, i.e. assume that $\Sigma^* < 1$. Then the stochastic process $\{\pi_t\}^*$ with the coefficient vector $w^*$ is stationary by Lemma 2, and therefore it satisfies the assumptions of Lemma 1. It can thus be predicted by

$$\pi_t^e = W^{**} \pi_t$$

where the polynomial $W^{**}$ is such that $w^* \leq w^{**}$ and $\Sigma^* < \Sigma^{**}$. But then the sequence $\{\ldots, \Sigma^{\nu}, \ldots\}$ cannot converge monotonically to $\Sigma^*$. This means that our assumption cannot be true, i.e. $\Sigma^*$ must be equal to unity.

We thus have a sequence of coefficient vectors $w^{\nu}$, and to each of these vectors is associated an inflationary process $\{\pi\}^{\nu}$, given by

$$\pi_t = f + W^{\nu} \pi_t + \varepsilon_t.$$

For finite $\nu$, all these processes satisfy the assumptions of Lemma 2 and are thus stationary. But what about the process associated with the limiting vector $w^*$? This process, $\{\pi\}^*$, is given by

$$\pi_t = f + W^* \pi_t + \varepsilon_t.$$

The polynomial $W^*$ is such that all its coefficients $w_j^*$ are non-negative. Further, the coefficients sum to unity. The process $\{\pi_t\}^*$ thus satisfies the assumptions of Lemma 3 and is therefore non-stationary, with all its characteristic roots on or outside the unit circle in the complex plane (with $|\zeta_i| = 1$ for at least one i). Such processes have realizations that look very similar to those of an ordinary random walk, with drift,

$$\pi_t = m + \pi_{t-1} + \varepsilon_t.$$

For that reason we label them "random-walk-like".

We have shown that the sequence of coefficient vectors $\{\ldots, w^\nu, \ldots\}$ converges to a vector $w^*$, and that the stochastic process associated with $w^*$ is non-stationary, random-walk-like with drift. Can we thereby say that the sequence of stochastic processes

$$\left\{\ldots, \{\pi_t\}^{\nu-1}, \{\pi_t\}^{\nu}, \{\pi_t\}^{\nu+1}, \ldots\right\}$$

converges to the random-walk-like process $\{\pi_t\}^*$? This is not quite certain; convergence of vectors $w^\nu$ in $R^n$ is a fairly straightforward thing, while convergence of random variables is a more tricky matter. The question might seem academic, but it has an important economic interpretation. If we can say that the sequence of processes $\{\pi_t\}^\nu$ converges to a random--walk-like process, this means that as $\nu$ increases, the processes $\{\pi_t\}^\nu$ become *more and more similar to a random-walk--like process*. If we can only show that the coefficient vectors $w^\nu$ converge, but not that the processes $\{\pi_t\}^\nu$ converge, we cannot say that the processes become more and more similar to the $\{\pi_t\}^*$ process. We can of course characterize the process $\{\pi_t\}^*$ as random-walk-like, but since we do not know whether we will ever reach $\{\pi_t\}^*$, such a knowledge is perhaps not entirely interesting for economic policy.

By convergence of stochastic processes, we will in this context mean *convergence in distribution.*[4] Let $\{\pi_t\}^\nu$ be a sequence of stochastic processes with distribution functions

$$F^\nu(x) \equiv \mathrm{Prob}(\pi_{t_1} \leq x_{t_1}, \pi_{t_2} \leq x_{t_2}, \ldots).$$

Then the sequence of processes $\{\pi_t\}^\nu$ is said to converge in distribution to $\{\pi_t\}^*$ if there exists a process $\{\pi_t\}^*$ with a distribution function $F^*(x)$ such that

$$\lim_{\nu\to\infty} F^\nu(x) = F^*(x)$$

for all x at which $F^*$ is continuous.

Now, convergence in $R^n$ of the coefficient vectors $w^\nu$ implies convergence in distribution of the processes $\{\pi_t\}^*$ if the distribution function $F^\nu$ is continuous in $w^\nu$, and we will demonstrate that this is the case.

Let us regard an AR(n) process

$$\pi_t = f + \sum_{j=1}^{n} w_j \pi_{t-j} + \varepsilon_t.$$

If the disturbances $\varepsilon_t$ are normally distributed, the process $\{\pi_t\}$ will be Gaussian, i.e. its distribution function will be completely characterized by the mean and the covariance matrix of the process. Thus, if the mean and the autocovariances are continuous in the coefficients $w_j$, then the distribution function will also be continuous in the coefficients.

Now it is easier to deal with the process when written on the equivalent MA form. We thus have

$$\pi_t = m + \sum_{j=1}^{\infty} \beta_j \varepsilon_{t-j} + \varepsilon_t.$$

---

[4] For an account of different aspects of convergence of random variables, see e.g. Feller (1971, chapter 8) or, more elementary, Karlin and Taylor (1975, chapter 1).

Since the AR coefficients are continuous in the MA coefficients, it is sufficient to show that the mean and the autocovariances are continuous in the $\beta_j$. Assume we have a realization of the process $\bar{\pi}_n$ (where a bar indicates a realized value). We then have the realizations of the disturbances $\bar{\varepsilon}_n, \bar{\varepsilon}_{n-1}, \bar{\varepsilon}_{n-2}, \ldots$

Let us denote the mean $E[\pi_t | \bar{\varepsilon}_n, \bar{\varepsilon}_{n-1}, \ldots]$ by $\mu_t$. Thus

$$\mu_t = m + \sum_{j=o}^{\infty} \beta_{t-n+j} \bar{\varepsilon}_{n-j}$$

and similarly, for $s \geq t$,

$$\mu_s = m + \sum_{j=o}^{\infty} \beta_{s-n+j} \bar{\varepsilon}_{n-j}$$

where the sums are finite by the assumption that $\bar{\pi}_n$ is finite. Thus the mean $\mu_t$ is continuous in $\beta_j$ for all t, as was to be shown. For the covariance we have

$$\mathrm{Cov}[\pi_t, \pi_s | \bar{\varepsilon}_n, \bar{\varepsilon}_{n-1}, \ldots] \equiv E[(\pi_t - \mu_t)(\pi_s - \mu_s) | \bar{\varepsilon}_n, \bar{\varepsilon}_{n-1}, \ldots] =$$

$$= m + m \sum_{j=o}^{\infty} \beta_{s-n+j} \bar{\varepsilon}_{n-j} + \sigma_\varepsilon \sum_{j=1}^{t-n-1} \beta_j \beta_{s-t+j} +$$

$$+ \sum_{i=o}^{\infty} \left( \beta_{t-n+i} \bar{\varepsilon}_{n-i} \sum_{j=o}^{\infty} \beta_{s-n-j} \bar{\varepsilon}_{n-j} \right).$$

Thus the autocovariances are also continuous in the $\beta_j$ coefficients. Since the AR coefficients are continuous in the MA coefficients, we also have that the mean and the covariances are continuous in $w_j$. The distribution function is thereby continuous in the $w_j$ coefficients, from which follows our desired result: the sequence of stochastic processes converges in distribution to the process $\{\pi_t\}^*$.

The rest of this section is confined to the proofs of the above Lemmas.

*Proof of Lemma 1:*

The optimal prediction is given by the orthogonality conditions

$$Rw = r. \tag{5.8}$$

Since the second moment $r_\tau$ is related to the covariance $\gamma_\tau$ by

$$r_\tau = \gamma_\tau + \mu^2 \qquad \tau = 0, 1, 2, \ldots$$

(where $\mu \equiv E[\pi_t] = \frac{m}{1-\Sigma\alpha_j}$) , we have

$$R \equiv \Gamma + \mu^2 A$$

and

$$r = \gamma + \mu^2 a$$

where A is the unit matrix

$$A \equiv \begin{pmatrix} 1 & 1 & \cdot & \cdot & \cdot & 1 \\ 1 & 1 & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & 1 \end{pmatrix}$$

and

$$a \equiv (1, 1, \ldots, 1)^T.$$

Therefore the system (5.8) can be written

$$(\Gamma + \mu^2 A)w = \gamma + \mu^2 a. \tag{5.9}$$

We have the Yule-Walker equations

$$\Gamma\alpha = \gamma$$

which, subtracted from (5.9) yields

$$\Gamma(w-\alpha) = \mu^2(a-Aw). \tag{5.10}$$

Note that (a-Aw) is the column vector

$$(a-Aw) = \begin{pmatrix} 1 - \Sigma w_j \\ 1 - \Sigma w_j \\ \cdot \\ \cdot \\ \cdot \\ 1 - \Sigma w_j \end{pmatrix}$$

For a stationary process, $\Sigma w_j < 1$ as has been shown in the previous chapter. Thus $1 - \Sigma w_j$ is a positive number, say k, and (a-Aw) can conveniently be written as $k \cdot a$. Equations (5.10) thus yield

$$w - \alpha = \mu^2 k \cdot \Gamma^{-1} a \tag{5.11}$$

To prove the *second* part of Lemma 1, we premultiply (5.11) by $a^T$ to obtain

$$a^T w - a^T \alpha = \mu^2 k \cdot a^T \Gamma^{-1} a.$$

Covariance matrices are always positive definite; the inverse of a positive definite matrix is also positive definite. Hence the quadratic form $a^T\Gamma^{-1}a$ is positive, and since k is positive and $\mu^2 \geq 0$ we have that

$$\Sigma w_j \geq \Sigma \alpha_j$$

with strict inequality if $\mu \neq 0$, i.e. if $f \neq 0$.

To prove the *first* part of Lemma 1 is more difficult. We see from (5.11) that $w-\alpha \geq 0$ if $\Gamma^{-1}a \geq 0$. This property can be analytically proved for AR(n) processes with $n \leq 2$; for such processes we have

$$\Gamma = \begin{pmatrix} \gamma_o & \gamma_1 \\ \gamma_1 & \gamma_o \end{pmatrix}$$

and

$$\Gamma^{-1} = \frac{1}{\gamma_o^2 - \gamma_1^2} \begin{pmatrix} \gamma_o & -\gamma_1 \\ -\gamma_1 & \gamma_o \end{pmatrix} .$$

Thus

$$\Gamma^{-1}\, a = \frac{1}{\gamma_o^2 - \gamma_1^2} \begin{pmatrix} \gamma_o - \gamma_1 \\ \gamma_o - \gamma_1 \end{pmatrix}$$

which is always strictly positive since $\gamma_o > |\gamma_\tau|$, $\tau \neq 0$. I have not succeeded, however, to obtain any analytical proof for processes of higher order, but have made numerical calculations to support the conjecture. I have let a computer generate n+2 non-negative random numbers $f$, $\alpha_1$, $\alpha_2$, ... $\alpha_n$ and $\sigma_\varepsilon^2$.Then the covariance matrix of the thereby obtained process

$$\pi_t = f + \sum_{j=1}^{n} \alpha_j\, \pi_{t-j} + \varepsilon_t$$

has been computed and inverted, to check whether $\Gamma^{-1}a \geq 0$. In this way, 20.000 processes of different orders have been randomly generated and checked, and no counterexample of the conjecture has been found. A more detailed description of the procedure can be found in Persson (1979). Although such a large sample from the (infinite) population of AR processes is quite convincing, it is clear that for processes of orders higher than 2, the second part of Lemma 1 is still but a conjecture.

*Proof of Lemma 2:*

The AR(n) process is stationary if and only if the roots of its characteristic equation lie outside the unit circle in the complex plane. We should thus demonstrate that if $w_j \geq 0$, all j, and $\Sigma\alpha_j < 1$, then the equation

$$1 - \sum_{j=1}^{n} w_j \; z^j = 0$$

has all its roots outside the unit circle. This can be proved by contradiction. Assume there exists some root $\zeta_k$ such that $|\zeta_k| \leq 1$. We then have

$$1 = |\Sigma w_j \; \zeta_k^j| \leq \Sigma w_j |\zeta_k^j| \leq \Sigma w_j$$

where the first of the inequalities follows from the definition of a norm in linear space, and the second follows from the Cauchy-Schwarz Inequality. But this contradicts the assumption that $\Sigma w_j < 1$. Thus, no such $\zeta_k$ exists.

Q.E.D.

*Proof of Lemma 3:*

Setting $\zeta_i = 1$ yields $1 - \Sigma w_j = 0$, which proves that $|\zeta_i| = 1$ for at least one i. To prove the first part of the Lemma, we assume that there exists some root $\zeta_k$ such that $|\zeta_k| < 1$. We then have

$$1 = |\Sigma w_j \; \zeta_k^j| \leq \Sigma w_j |\zeta_k^j| < \Sigma w_j .$$

This contradicts the assumption that $\Sigma w_j = 1$. Thus no such $\zeta_k$ can exist.

Q.E.D.

## 5.4 MULTI-PERIOD FORECASTING

In the preceding sections we have studied the effects on a simple macro model of the adaptive expectations formula

$$\pi^e_{t+\theta} = \sum_{j=1}^{n} w_j \, \pi_{t-j} \tag{5.12}$$

for the particular case of $\theta = 0$, i.e. for a one-period forecasting horizon. There is however no reason why the length of the (implicit or explicit) contracts that govern prices and inflation should be exactly one period. In fact, it seems even likely that in reality multi-period contracts are more common than one-period contracts. In any case, since by definition no contract can be shorter than one time period, but can easily be longer, the *average* contract is longer than one time period. For an exhaustive analysis of the multi-period expectations-augmented Phillips curve,[5] we should of course formulate it including several terms $\pi^e_{t+\theta}$ with different horizons $\theta$. This, however, becomes exceedingly complicated; we thus let all the different contract lengths be collapsed into one. Even such a simplification is not free from difficulties. The multi-period Phillips curve can be written alternatively

$$\pi_t = f(u_t) + {}_t\pi^e_{t+\theta} + \varepsilon_t \tag{5.13a}$$

or

$$\pi_t = f(u_t) + {}_{t-\theta}\pi^e_t + \varepsilon_t . \tag{5.13b}$$

The formulation (5.13a) means that inflation at time t is given by the unemployment rate at time t and by the rate of inflation forecasted at time t to prevail at time t+$\theta$ (the first subscript of $\pi^e$ indicates the date at which the forecast is made). Such an interpretation can be justified if

---

[5] For a treatment of multi-period contracts in another type of macro model, see Fischer (1977).

the Phillips relation mirrors agents trying to compensate themselves today for expected price changes $\theta$ periods in the future, thereby setting prices at time t as a function of prices expected to prevail at time $t+\theta$. The second formulation, (5.13b), can be justified as describing an economy where the agents entered long-term contracts $\theta$ periods ago, thereby determining which prices should rule today.

In the previous section, with $\theta = 0$, we disregarded this ambiguity in the expectations term. The formula

$$\pi_t = f(u_t) + \sum_{j=1}^{n} w_j \pi_{t-j} \tag{5.14}$$

*could* mean that the world is like in (5.13a); agents make guesses at time t of the inflation rate prevailing at that time, the only information available being the inflation rates through time t-1. *Also*, it could mean that the agents at time t-1, with the information available then including $\pi_{t-1}$, make forecasts of tomorrow's inflation $\pi_t$. Although these two interpretations of (5.14) differ widely, both on the conceptual level and concerning the assumptions about the availability of recent economic data, either of them could perhaps be accepted as "reasonable". When $\theta > 0$, however, the arbitrariness in the interpretation of the expectations term becomes only too obvious, and we should therefore analyze both the Phillips curve (5.13a),

$$\pi_t = f(u_t) + \sum_{j=1}^{n} w_j \pi_{t-j} + \varepsilon_t \tag{5.15a}$$

(where the $w_j$ weights are such as to give an optimal adaptive forecast of $\pi_{t+\theta}$), and the corresponding (4.13b) version

$$\pi_t = f(u_t) + \sum_{j=1}^{n} w_j \pi_{t-\theta-j} + \varepsilon_t . \tag{5.15b}$$

Note that they differ widely; if $f(u_t)$ is pegged to a constant f, (5.15a) implies that $\pi_t$ follows an AR(n) process, while (5.15b) implies an AR($n+\theta$) process.

In Chapter 4 above we saw that the multi-period case does not lend itself to very straightforward analysis. The optimal weight sum is not necessarily less than unity, and it could be both increasing and decreasing in $\theta$. Analytical results concerning the existence or non-existence of a natural rate are therefore not to be expected. Instead we will falsify the NRH by a very simple numerical example.

Assume that the length of the agents' memory, i.e. the length n of the adaption formula, is unity. We then have the two Phillips curves

$$\pi_t = f(u_t) + w\pi_{t-1} + \varepsilon_t \tag{5.16a}$$

and

$$\pi_t = f(u_t) + w\pi_{t-\theta-1} + \varepsilon_t. \tag{5.16b}$$

We start with the first one, assume that the authorities peg unemployment to u so that $f_t = f$, and start at a stationary, natural rate equilibrium,

$$\pi_t = \varepsilon_t,$$

where the optimal weight w = 0. If after some time f is changed to a value $\neq 0$, we have the process $\{\pi_t\}^1$ defined by

$$\pi_t = f + \varepsilon_t.$$

Soon enough the agents start to form an optimal expectation of this process,

$$\pi^e_{t+\theta} = w^1 \pi_{t-1}.$$

The $w^1$ coefficient is given by the orthogonality conditions

$$w^1 = r_{\theta+1}^{(1)} \;/\; r_o^{(1)} = \frac{f}{f + \sigma_\varepsilon^2}$$

where the superscript $^{(1)}$ of the second moment terms refers to the process $\{\pi_t\}^1$. We thus get a new process $\{\pi_t\}^2$:

$$\pi_t = f + w^1 \pi_{t-1} + \varepsilon_t.$$

This can be optimally predicted by a $w^2$ coefficient given by

$$w^2 = r_{\theta+1}^{(2)} \;/\; r_o^{(2)} = \frac{\left(1 - (w^1)^2\right)\left(\frac{f^2}{1-w^1} + w^1 r_\theta^{(2)}\right)}{f^2\left(1 + \frac{2w^1}{1-w^1}\right) + \sigma_\varepsilon^2}$$

In general the process $\{\pi_t\}^{\nu+1}$ will be described by a system of different equations

$$w^\nu = r_{\theta+1}^{(\nu)} \;/\; r_o^{(\nu)} \tag{5.17}$$

$$r_{\theta+1}^{(\nu)} = \frac{f^2}{1 - w^{\nu-1}} + w^{\nu-1} r_\theta^{(\nu)} \tag{5.18}$$

with the initial conditions

$$w^1 = \frac{f}{f + \sigma_\varepsilon^2} \tag{5.19}$$

$$r_o^{(\nu)} = \frac{f^2\left(1 + \frac{2w^{\nu-1}}{1 - w^{\nu-1}}\right) + \sigma_\varepsilon^2}{1 - \left(w^{\nu-1}\right)^2} \tag{5.20}$$

This system is not very convenient, but it can be solved numerically. Whether or not the sequence $\{\ldots, w^{(\nu)}, \ldots\}$ converges to a value indicating a stationary inflation process, i.e. to a $w^*$ which is less than unity in absolute value, depends on the parameters f, $\theta$ and $\sigma_\varepsilon^2$. In Table 5.1 are displayed computations for different sets of parameters, and we see that sometimes the equilibrium process $\{\pi_t\}^*$ will actually be stationary:

Table 5.1: Values of w* for a first-order adaptation formula with $\sigma_\varepsilon^2 = 5$.
(N. stands for "non-stationary".)

| $\theta$ \ f | 0.5 | 0.75 | 1.0 | 1.25 | 1.5 |
|---|---|---|---|---|---|
| 0 | N. | N. | N. | N. | N. |
| 1 | 0.0561 | 0.1601 | N. | N. | N. |
| 2 | 0.0528 | 0.1296 | 0.2929 | N. | N. |
| 3 | 0.0526 | 0.1271 | 0.2573 | N. | N. |
| 4 | 0.0526 | 0.1268 | 0.2517 | N. | N. |
| 5 | 0.0526 | 0.1268 | 0.2504 | 0.4967 | N. |
| 10 | 0.0526 | 0.1268 | 0.2500 | 0.4550 | N. |
| 15 | 0.0526 | 0.1268 | 0.2500 | 0.4545 | N. |
| 20 | 0.0526 | 0.1268 | 0.2500 | 0.4545 | N. |
| 25 | 0.0526 | 0.1268 | 0.2500 | 0.4545 | N. |
| 50 | 0.0526 | 0.1268 | 0.2500 | 0.4545 | 0.8155 |

That is, if the agents make two-periods-ahead forecasts (corresponding to $\theta = 1$), and the authorities try to peg unemployment at a value u such that $f(u) \equiv f = 1.25$, then this will turn out to be impossible in the long run, since inflation will ultimately explode. But if the authorities are less ambitious, and peg f at, say, $f = 0.75$ (and the agents still form two-periods-ahead forecasts), inflation will ultimately stabilize at a stationary process

$$\pi_t = 0.75 + 0.1601\ \pi_{t-1} + \varepsilon_t.$$

Similarly, if the agents enter such long-term contracts as corresponding to $\theta = 50$ (which is hardly realistic; we mention it here for illustrative purposes only), the authorities can be really ambitious. Pegging u at such a low value that $f(u) = 1.5$ yields eventually the stationary inflation process

$$\pi_t = 1.5 + 0.8155\ \pi_{t-1} + \varepsilon_t.$$

This means that there exists some region $f_{min} \leq f \leq f_{max}$ within which the authorities can affect unemployment and obtain an equilibrium process $\{\pi_t\}^*$ which is stationary. The long run Phillips curve is thus not vertical for this interval, which depends on the exogenous parameters $\theta$ and $\sigma_\varepsilon^2$:

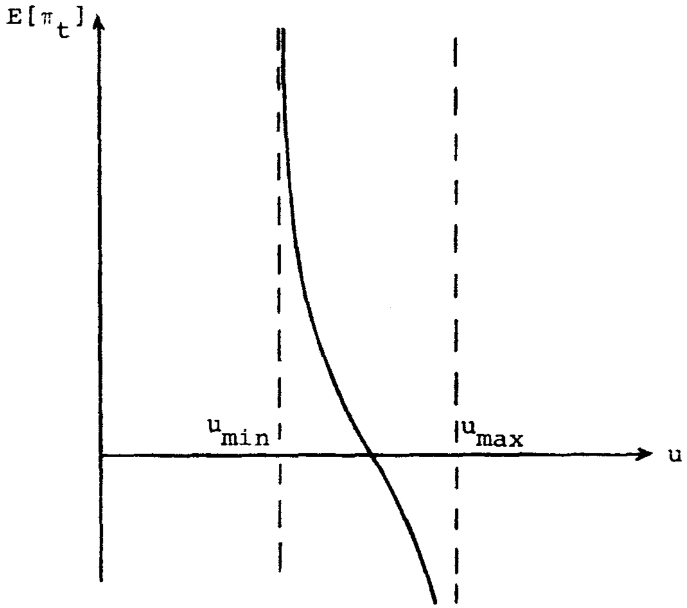


Figure 5.2: The long-run Phillips curve with multi-period contracts.

Let us finally go back to our equations (5.16). The figures in Table 5.1 refer to the image of the world reflected in (5.16a), and we will now trace out the consequences of (5.16b). We start in the same way as before, with a stationary, natural rate equilibrium $\pi_t = \varepsilon_t$ and proceed exactly as before. The expressions will be fairly similar, but not quite. The $w^\nu$ weight will be given by an equation which looks the same as (5.17), namely

$$w^\nu = r_{\theta+1}^{(\nu)} \;/\; r_o^{(\nu)}$$

However, since the Phillips equation (5.16b) is now an AR($\theta$+1) process, $r_{\theta+1}^{\nu}$ will not be the same as before; the expression corresponding to (5.18) will instead be

$$r_{\theta+1}^{(\nu)} = \frac{f^2}{1 - w^{\nu-1}} + w^{\nu-1}\; r_o^{(\nu)}$$

with $w^{\nu-1}$ and $w^1$ given by expressions exactly similar to (5.19) and (5.20), respectively. Substituting yields

$$w^\nu = \frac{f^2}{\left(1 - w^{\nu-1}\right) r_o^{(\nu)}} + w^{\nu-1} \tag{5.21}$$

Since $w^{\nu-1} = r_{\theta+1}^{\nu-1} \;/\; r_o^{\nu-1} < 0$, we see that the first term on the right hand side will always be positive if $f \neq 0$. Hence the differential equation (5.21) will explode, i.e. inflation will explode too. Thus, if the world of multi-period forecasting works as implied by the Phillips curve (5.16b), the NRH holds *regardless of the length of the forecasting horizon.*

## 5.5 CONCLUDING COMMENTS

In this chapter we have demonstrated that if the agents form one-period optimal adaptive forecasts, then the NRH holds for the expectations-augmented Phillips curve model. The result depends basically on the crucial assumption that the agents actually *form* adaptive forecasts. This assumption was the point of departure for our whole analysis, and although this assumption is standard in economic and econometric literature (which makes it a justified object of study) we should perhaps briefly discuss whether it is a realistic and a reasonable assumption.

Assume that inflation is given by one of the $\{\pi_t\}^{\nu+1}$ processes discussed in Section 5.3 above,

$$\pi_t = f + \sum_{j=1}^{n} w_j^{\nu} \pi_{t-j} + \varepsilon_t. \tag{5.21}$$

Why should then people be so irrational that they form expectations according to

$$\pi_t = \sum_{j=1}^{n} w_j^{\nu+1} \pi_{t-j} \tag{5.22}$$

instead of just taking the conditional expectation of (5.21) as their forecast? The proper answer to this is, I think, the following: It is difficult to model expectations in an operational way. We have only two mechanisms that are operational, the adaptive expectations (5.22) and the rational expectations

$$\pi_t^e = E[\pi_t | \pi_{t-1}, \pi_{t-2}, \ldots]. \tag{5.23}$$

We know that it is irrational to form expectations according to (5.22) if inflation is given by (5.21). On the other hand, we also know that people are not well-informed enough to form expectations according to (5.23). What we can do as economists is to study all existing formulas, neither of them being entirely realistic, and see whether the NRH is *robust* in the sense that our models have the same natural rate properties regardless of which expectations formula is applied. If this is the case, we are allowed to form opinions of whether the NRH holds in the real world. And while Chapters 4 and 5 have dealt with the adaptive formula, we are now going to study the rational formula in Chapters 6 and 7.

There are several other assumptions that do not seem altogether realistic, the most important being perhaps the one we discussed on page 101 above, namely that we choose to disregard the transition time between two processes $\{\pi_t\}^{\nu}$ and $\{\pi_t\}^{\nu+1}$. The conclusion of the one-period forecasting model, that the inflationary process becomes more and more similar to a random walk with drift (convergence in distribution) might however hold in reality in a less precise form; if f is kept above the natural value zero, both the mean and the variance in the inflationary process will tend to increase over time.

Finally, we should mention a few words about the multi-period results of Section 5.4. These are important since they show that the NRH is not quite robust; if the Phillips curve is formulated as (5.16a) there is a long-run trade-off between $E[\pi_t]$ and f. We have not defined the length of the time unit in our analysis; thus it is perhaps difficult to judge whether people actually enter multi-period contracts in reality, and even more difficult to say whether these contracts imply a Phillips curve according to (5.16a) or (5.16b). In empirical work the time unit is automatically defined by the availability of data; if monthly data are available, the time unit is a month, and if only quarterly data are available, it is a quarter. In a theoretical model it is reasonable that the time unit should be equal to the time it takes for the authorities to react, i.e. to the sum of the observation and reaction lags. Whatever this might be, it seems likely that the authorities have shorter observation lags than the individual agents, and even if the length of the decision lag is somewhat ambiguous, it does not seem entirely unrealistic to assume that the agents generally enter multi-period contracts (i.e. have less flexibility than the authorities). The falsification of the NRH that thereby follows from (5.16a) is thus of the same kind as Sargent's and Wallace's (1975) "information advantage for the monetary authority", which also in their model implies a falsification of the NRH.

# 6. RATIONAL EXPECTATIONS IN LOG-LINEAR MODELS

The natural rate hypothesis (NRH) says that real variables cannot be systematically affected by the nominal variables of the economy. This refers to all real variables, i.e. the rate of unemployment, the level of aggregate, real output, or the real rate of interest. In the preceding chapter we studied the NRH in a Phillips curve context, and therefore we concentrated our attention to the unemployment rate. In this chapter, dealing with rational expectations, we will conform to the tradition of the rational expectations literature and concentrate on the level of real output.

In this context the NRH can be formulated in a way analogous to the one in Chapter 5 above: Denote real output and the price level at time t by $y_t$ and $p_t$, respectively, and assume that they follow time series $\{\ldots, y_{t-1}, y_t, y_{t+1}, \ldots\}$ and $\{\ldots, p_{t-1}, p_{t+1}, \ldots\}$ with means $E[y_t]$ and $E[p_t]$. *The natural rate hypothesis is said to hold if $E[y_t]$ is independent of the time series $\{\ldots, p_{t-1}, p_t, p_{t+1}, \ldots\}$ and, in particular, if there is no trade-off between $E[y_t]$ and $E[p_t]$.*

The exact relation between $y_t$ and $p_t$, and between their means, depends on two crucial factors: the model chosen to represent the economy, and the way expectations are assumed to be formed in the model. Lucas has claimed (1972 a) that rational expectations are equivalent to the NRH. This seems to be true for some general equilibrium models where agents display maximizing behavior.[1] In econometric work, however, such models are seldom used, and instead we have to rely on simple *ad hoc* models like the expectations-augmented Phillips curve or more elaborate, but still *ad hoc*, models like the ones of Sargent (1973) and Sargent-Wallace (1975). In such models, the *aggregate supply equation* is utilized to analyze the relations between real output, the price level, and the expected price level:

$$Y_t = f(P_t - P_t^e) + \varepsilon_t \tag{6.1}$$

where $Y_t \equiv \log y_t$, $P_t \equiv \log p_t$ and $\varepsilon_t$ is normally distributed white noise. The purpose of this chapter is to show that for such models, rational expectations are *not* equivalent to the NRH. On the contrary, we will show that real output can be systematically affected by monetary policy even if expectations are formed rationally in the sense of Muth and Lucas.

## 6.1 MODELLING MONETARY POLICY

To trace out the effects of monetary policy, we have to state which are the policy parameters at the authorities' disposal. While we in Chapter 5 above had a fairly eclectic view as to how economic policy is performed, i.e. whether the authorities control $\{y_t\}$ or $\{p_t\}$, and whether the direction of causality goes from $\{y_t\}$ to $\{p_t\}$ or vice versa, the rational expectations literature is quite explicit on that point (cf. Chapter 3 above): the authorities control the $\{p_t\}$ time series - for example via the money supply - and this affects real output $\{y_t\}$. More specifically, we model economic policy by first stating what is sometimes called "the aggregate demand sche-

---

[1] Cf. Lucas (1972 b).

dule",[2] but what is rather to regard as *the price identity*, equating nominal supply and demand:

$$Y_t + P_t = X_t \tag{6.2}$$

where $X_t$ denotes the logarithm of aggregate, nominal demand. A "monetary policy rule" is then defined in the standard way, namely by the parameters of a time series

$$X_t = k + \sum_{i=1}^{n} \alpha_i X_{t-1} + \eta_t. \tag{6.3}$$

The problem is to find k, $\alpha_1, \ldots, \alpha_n$ and $\sigma_\eta^2$ of (6.3) such that the long-run average value of real output can be affected.[3] Since for each time series $\{\ldots, X_{t-1}, X_t, X_{t+1}, \ldots\}$ there exists a time series of logs of prices $\{\ldots, P_{t-1}, P_t, P_{t+1}, \ldots\}$ we can, for the present purpose, simplify matters and define a policy as the time series of the prices themselves, or of their logarithms.

We immediately see that if the supply function f(.) is nonlinear, then the average value $E[Y_t]$ could be changed by a policy which affected the variance of $(P_t-P_t^e)$.[4] Assume for example that f is quadratic:

$$Y_t = a + b(P_t-P_t^e) + c(P_t-P_t^e)^2 + \varepsilon_t. \tag{6.4}$$

The rational expectations are given by

---

[2] See e.g. Lucas (1972 a, p. 55).

[3] If the authorities cannot control nominal demand completely (as seems reasonable to assume) the role of $\sigma_\eta^2$ as a policy parameter is somewhat limited by the fact that it cannot be decreased below a certain minimum value.

[4] In Lipsey's 1960 article, a similar point is made on the curvature of the Phillips relation. There the variance referred to dispersion between different markets, but the argument naturally carries over to the time series context of this chapter.

$$P_t^e \equiv E[P_t | H_{t-1}]$$

where $H_{t-1}$ denotes past history through time t-1. If the authorities can control the money supply so that the nominal variable follows an autoregressive time series

$$P_t = \delta + \sum_{i=1}^{n} \beta_i P_{t-i} + v_t, \tag{6.5}$$

where $v_t$ is white noise with variance $\sigma_v^2$, we can obtain the long-run average value of $Y_t$ by substituting (6.5) into (6.4) and take expectations; this yields

$$E[Y_t] = c\ \sigma_v^2 + a.$$

Thus, by changing $\sigma_v^2$ the authorities can affect $E[Y_t]$. If the supply function f(.) is concave, i.e. if $c < 0$, the authorities can maximize $E[Y_t]$ by minimizing the stochastic disturbance. Since the white noise is generally considered to be uncontrollable, there probably exists some minimum level of $\sigma_v^2$ below which the economy cannot reach. Thus, there exists some maximal attainable level of $E[Y_t]$, and this is the closest correspondence to the natural rate hypothesis we can obtain in this simple model.

A somewhat more disturbing feature accrues if f(.) happens to be convex. Then $c > 0$, $E[Y_t]$ increases with $\sigma_v^2$ as long as (6.4) is a good approximation of economic reality, and a highly (or even infinitely) erratic monetary policy would then be desirable. Whether f(.) is concave or convex remains to be investigated, empirically and theoretically; may it suffice here to emphasize the effects that can occur if the f(.) function has the "wrong" curvature.

Now, it is a fairly well-known fact that a supply function (6.1) which is nonlinear in its argument $(P_t - P_t^e)$ will not in general display natural rate properties[5], and we

[5] See for example Shiller (1978, p. 10).

should not go into these questions further. Instead, I will follow most writers dealing with rational expectations and assume that (6.1) is *linear*, i.e. of the form

$$Y_t = a + b(P_t - P_t^e) + \varepsilon_t \tag{6.6}$$

where a and b are positive constants.

If we set $P_t^e \equiv E[P_t|H_{t-1}]$ in (6.6) and take the unconditional expectation of both members, we get $E[Y_t] = a$. This is the "natural rate result" obtained by Lucas (1972 a, 1973), Sargent and Wallace (1975), Barro (1976), Fischer (1977), and numerous other writers who have dealt with a supply function of the form (6.6). However, I will claim that this does not imply natural rate properties for the above model; it only says that since (6.6) is linear in its arguments, $E[Y_t]$ cannot be affected by monetary policy. What society is interested in is *not* $E[Y_t] \equiv E[\log y_t]$, but $E[y_t]$. And the latter can be affected by monetary policy. In particular, I will show that a linear function (6.6) and a nonlinear transformation $Y_t \equiv \log y_t$ give the "wrong" curvature to the relation between $y_t$ and $p_t$, so that in models with the standard definition of rational expectations $E[y_t]$ will be an *increasing* function of $\mathrm{Var}[p_t]$.

## 6.2 THE LOGARITHMIC TRANSFORMATION

Assume that the economy can be sufficiently well described by the linear supply function (6.6). Now, "rational expectations" could mean two things. Either

$$P_t^e \equiv E[\log p_t|H_{t-1}] \tag{6.7}$$

where $H_{t-1}$ denotes past history through time t-1, or

$$P_t^e \equiv \log \left(E[p_t | H_{t-1}]\right). \tag{6.7'}$$

Both interpretations, regarded as behavioral relations, are equally plausible, and they seem to have equal support in the literature. They are equivalent if the model in which they are employed is *deterministic*; for *stochastic models*, however, they have different implications. In the following they will be denoted "The First Definition" and "The Second Definition", respectively, and it will be demonstrated that none of them is equivalent to the natural rate hypothesis.

### 6.2.1 The First Definition

Rational expectations defined according to (6.7) are the ones that are generally utilized in macro models, and the ones that underlie the natural rate properties derived from these models. Thus, Lucas (1972 a, p. 54-55) states that the relation

$$E[P_t - P_t^e \mid H_{t-1}] = 0,$$

introduced into the model as an additional axiom, implies that we have assumed expectations to be rational in the sense of Muth. This obviously means that $P_t^e$ should be defined according to (6.7).

Let us first look at the instantaneous effect on $y_t$ from a change in $p_t$, i.e. the effect before expectations have adjusted. From (6.6) we see that

$$\frac{\partial y_t}{\partial p_t} = b\,\frac{y_t}{p_t}.$$

The increase in $y_t$ resulting from an unexpected increase in the price level is greater the greater is the value of $y_t$, implying that at the peak of the business cycle, an extra monetary stimulus will have greater impact than at the bottom of the cycle. This fact, which has been pointed out by David Laidler (1975, p. 186), obviously indicates that there is something strange with the curvature of the logarithmic transformation. This refers only to the *short-run* effect of an instantaneous change in $y_t$, but it also carries over to the long-run properties we study in the natural rate context. Assume $p_t$ follows a stationary time series $\{\ldots, p_t, \ldots\}$ and define rational expectations according to (6.7). Taking the unconditional mean of (6.6) then yields

$$E[Y_t] = a$$

regardless of the time series $\{p_t\}$. This is the natural rate result one would expect, but notice that it only says that the mean of log $y_t$ is independent of $\{p_t\}$; it might well be that the average value of $y_t$ itself could be affected by the nominal variable. In fact, expanding $Y_t = \log y_t$ around the point $y_t = E[y_t] \equiv \bar{y}$ yields

$$\log y_t \approx \log \bar{y} + (y_t - \bar{y}) \frac{1}{\bar{y}} - \frac{1}{2!\bar{y}^2} (y_t - \bar{y})^2 \qquad (6.8)$$

where the higher-order terms are disregarded. Taking the mathematical expectation of (6.8) yields, since $E[\log y_t] = a$,

$$\bar{y}(\log \bar{y} - a) \approx \frac{1}{2} \mathrm{Var}[y_t].$$

Denoting the function $\bar{y}(\log \bar{y} - a)$ by $\phi(\bar{y})$, we have

$$\bar{y} \approx \phi^{-1}\left(\frac{1}{2} \mathrm{Var}[y_t]\right).$$

The $\phi$ function is increasing, and thus $\phi^{-1}$ is increasing too. $E[y_t]$ is hence an increasing function of $\mathrm{Var}[y_t]$. The consequence of this is that if the authorities employ a highly erratic monetary policy, with the variance of the white noise term in (6.3) going to infinity, then the average level of real output will increase without bound - at least as long as (6.6) and the linearization (6.8) can be regarded as a reasonable approximation to economic reality. If it had been the other way around, that is, if $\phi^{-1}$ had been a *decreasing* function instead, the model would have made more sense. Then $E[y_t]$ would have been a decreasing function of $\sigma_\eta^2$, and since the white noise is generally regarded as reflecting some uncontrollable disturbances to the economy, one could argue that $\sigma_\eta^2$ could not be reduced below some minimum value. Thus, $E[y_t]$ would have been bounded from above at a value which could perhaps be said to be the "natural level" of real output. Since this is not the case, however, and since there certainly is no upper bound on $\sigma_\eta^2$, the average value of real output can be increased *ad infinitum* in this kind of model.

In fact, this is quite a strong result, which relates to a much wider class of model formulations than the one above. Thus, it holds for all models of the form (6.6) where

(i) $Y_t$ is any concave function of $y_t$, and

(ii) $P_t^e$ is any unbiased expectations operator of $P_t$.

An unbiased expectations operator is an operator $Z_t^e$ of $Z_t$ such that $E[Z_t^e] = E[Z_t]$, a property which the rational expectations (6.7) share with for instance the adaptive expectation

$$Z_t^e = \sum_{j=1}^{n} w_j Z_{t-j}, \quad \Sigma w_j = 1.$$

### 6.2.2 The Second Definition

We see thus that if $P_t^e$ is defined according to (6.7), $E[Y_t] \equiv E[\log y_t]$ is independent of monetary policy, but the interesting variable, $E[y_t]$, is not. If on the other hand

$P_t^e$ is defined according to (6.7'), not even $E[Y_t]$, and certainly not $E[y_t]$, is independent of monetary policy. A question that naturally arises is whether (6.7') is a reasonable definition, and the answer seems to be that it might have more firmly established microfoundations than its counterpart (6.7). If, for example, an aggregate supply function is to be derived from a model of maximizing behavior where the agents (i.e. the workers) speculate over time, such as the Lucas-Rapping (1969) model, we obviously obtain (after some rather restrictive assumption) a supply function where the price expectations should be defined as $P_t^e \equiv \log \left(E[p_t | H_{t-1}]\right)$. This is also recognized by Lucas (1972 a), who defines $P_t^e$ as "the log of an index of expected future prices".[6]

Another way of justifying an expectations term like (6.7') is by a model like the one of Laidler (1978), where firms are speculating over markets. If $y_{it}$ is the output of the i:th firm at time t, $p_{it}$ is the price of that firm's output, and $p_{it}^e$ is the general price level as perceived by the firm, we can formulate a relation which states that firm i increases its output when it perceives its own relative price to be high:

$$y_{it} = f_i \left(\frac{p_{it}}{p_{it}^e}\right), \qquad f_i' > 0.$$

Given some rather restrictive assumptions[7] we can sum over the firms to obtain

$$y_t = F \left(\frac{p_t}{p_t^e}\right)$$

which, if F is log-linear, can be written

$$\log y_t = a + b(\log p_t - \log p_t^e).$$

---

6 Lucas (1972 a, p. 52).

7 For example, that all firms have identical reaction functions, and that all firms have the same perceptions of the general price level $p_t^e$.

We thus have a supply function of *the log of the expectation*, that is $p_t^e$ should be defined according to (6.7'). Obviously, this model does not have the time series approach which characterizes our earlier treatment of the supply function; $p_t^e$ refers in this model to the general price level, i.e. to an expectation *over markets*, while our earlier $p_t^e$, and the price expectation in the Lucas-Rapping model, refers to a forecast of *tomorrow's prices*. Nevertheless, it shows that quite reasonable micro models yield supply functions that imply a definition of $p_t^e$ similar to (6.7'), rather than to (6.7). This points at an ambiguity in the rational expectations models hitherto employed, and I will show that the non-existence of a natural rate of unemployment is robust with respect to this ambiguity.

Assume again that $p_t$ follows a stationary time series $\{\ldots, p_t, \ldots\}$, and let us for short introduce the notation $E[p_t] \equiv \mu$ and $E[p_t | H_{t-1}] \equiv m_t$. We have that $E[m_t] = \mu$ since $E[.|\ldots]$ is an unbiased operator. We substitute (6.7') into (6.6), take the unconditional expectation, and have

$$E[Y_t] = a + b(E[\log p_t] - E[\log m_t]) \qquad (6.9)$$

Expanding $\log m_t$ around the point $m_t = \mu$ yields

$$\log m_t = \log \mu + (m_t-\mu)\frac{1}{\mu} - \frac{1}{2!}(m_t-\mu)^2\frac{1}{\mu^2} + \\ + \frac{1}{3!}(m_t-\mu)^3\frac{1}{\mu^4} - \ldots \qquad (6.10)$$

Assuming that $\{\ldots, p_t, \ldots\}$ (and hence $\{\ldots, m_t, \ldots\}$) is a Gaussian process, all odd moments $M_k(m) \equiv E[(m_t-\mu)^k]$, $k = 1, 3, 5, \ldots$, vanish. Taking the mathematical expectation of (6.10) thus gives

$$E[\log m_t] = \log \mu - \frac{1}{2!\mu^2} M_2(m) - \frac{1}{4!\mu^8} M_4(m) - \ldots \qquad (6.11)$$

By the same reasoning, we can get an expression for $E[\log p_t]$:

$$E[\log p_t] = \log \mu - \frac{1}{2!\mu^2} M_2(p) - \frac{1}{4!\mu^8} M_4(p) - \ldots \quad (6.12)$$

Now, the even moments of the normal distribution have the property that, for given r, $M_r(.)$ is an increasing function of Var[.].[8] Since $Var[m_t] < Var[p_t]$, $M_r(m) - M_r(p)$ is a negative number for all r = 2, 4, 6, ... Further, if $Var[p_t]$ is held constant,

$$\frac{1}{\mu^k r!}\left(M_r(m) - M_r(p)\right)$$

is an increasing function[9] of $\mu \equiv E[p_t]$, which means that we have a trade-off between $E[Y_t]$ and $E[p_t]$. Since $Y_t \equiv \log y_t$, $Y_t$ could be linearized around the point $y_t = E[y_t]$, in the same way as $p_t$ and $m_t$, to yield

$$E[Y_t] = \log E[y_t] - \frac{1}{2!(E[y_t])^2} M_2(y) - \ldots$$

Thus, keeping the variance constant, we actually have a long-run trade-off between $E[y_t]$ and $E[p_t]$.

## 6.3 CONCLUDING COMMENTS

In deterministic models there is no ambiguity regarding how to model rational expectations, and for such models rational expectations are equivalent to the natural rate hypothesis. However, since we have only imperfect knowledge of the ultimate cause of all changes in the economic system, it seems appropriate to analyze the system by means of stochastic models. For such models, as usually formulated, the equi-

---

[8] Cf. Kendall and Stuart (1963, p. 60).

[9] Provided $\mu > 1$.

valence between rational expectations and the natural rate hypothesis does not hold.

The main point in this paper has been to show that although the widely used supply function (6.6) states $Y_t$ as a linear function of $P_t$, it implies that $y_t$, expressed as a function of $p_t$, has the "wrong" curvature. This means that if rational expectations are defined according to (6.7), $E[y_t]$ will be an *increasing* function of $Var[p_t]$, a result which is hardly in conformity with the general view that monetary policy should be aimed at minimizing the variance in $p_t$.

Two qualifications should be made: First, a supply equation like (6.6) is only to be regarded as a local approximation of the actual economy; it is not plausible that $E[y_t]$ will increase to infinity with $Var[p_t]$. But it is serious enough, when judging of the appropriateness of (6.6) as a description of the economic system, that it has the wrong curvature even in the limited range where it should be a good approximation to reality.

The second qualification concerns the appropriateness of *any* stable supply function (6.1) - linear or non-linear - in an economy where the authorities change between different policy régimes. This argument is similar to the one put forward by Lucas (1976 a), and is the one I am referring to at the beginning of this chapter, when discussing general equilibrium models that might display natural rate properties, as contrasted to *ad hoc* models which we have to rely on for practical purposes. In particular, it might be that for a proper derivation of a linear supply function like (6.6), the parameters turn out to be functions of the time series $\{..., p_t, ...\}$ announced by the authorities. Such an approach is taken by for example Lucas (1973) and Barro (1976), where the coefficient a is assumed to vary inversely with $Var[p_t]$.

However, the way this dependence is modelled in these papers is not sufficient to rule out the connections between $E[y_t]$ and the time series $\{..., p_t, ...\}$, the only way to completely rule out such a dependence being to build such a complicated model that we have in practice constructed a full-scale general equilibrium model.

The conclusion is thus that the way aggregate supply is generally represented in econometric models, i.e. by a log-linear function (6.6), is inappropriate, since it implies that the average level of real output increases monotonically with the variance in the money supply rule.

# 7. CERTAINTY EQUIVALENCE AND THE RATIONALE FOR RATIONAL EXPECTATIONS

Rational expectations have been critized from several different points of view. The *first* kind of critique, and the most fundamental one, concerns the assumption that economic agents are well-informed enough to form forecasts that are equal to the true, mathematical expectation of the variables in question. This is the kind of critique put forth by e.g. Shiller (1978) and B. Friedman (1979),[1] and we will not deal with it in the present chapter.

The *second* kind of criticism of the rational expectations is specifically directed towards a limited group of models, the *ad hoc* models of e.g. Lucas (1972 b, 1973), Sargent (1973) and Sargent and Wallace (1975). It says that these models do not rely on sound micro-foundations, i.e. they are not correctly derived from an assumption of maximizing behavior. This is the kind of criticism put forward by Fair (1978), and it does not really aim at the concept of rational expectations *as such*, but rather at the particular models in which rational expectations are utilized.

The *third* kind of criticism is closely related to the second one. It points out that, whether or not the ad hoc models referred to above rely on maximizing behavior, they

[1] This type of critique has been met by e.g. Barro and Fischer (1976, p. 163).

display properties which are counterintuitive to macroeconomic thinking, and thus there might be something wrong with them. In particular, it turns out that for such models the average value of real output is an *increasing* function of the variance in the money supply rule. This was demonstrated in the preceding chapter, and it is obvious that it is closely related to the second type of criticism just mentioned; if the models do not have solid micro-foundations, it is hardly surprising that they display strange, or even "wrong", properties.

In the present chapter we will pursue the second, and to some extent the third, line of criticism by investigating some of the micro-foundations of the ad hoc models. In particular, we will study the use of certainty equivalents in one of their essential features, namely the Lucas supply equation

$$Y_t = a + b(P_t - P_t^e). \tag{7.1}$$

As noted earlier, the origins of this function are somewhat obscure, but according to Lucas (1972 b, p. 52) it can be derived from Lucas' and Rapping's (1969) two-period consumer decision model, where the consumer has four decision variables: consumption of goods in period one, labor supply in period one, consumption in period two, and labor supply in period two. The agent thus maximizes expected utility

$$\max_{c_1, c_2, \ell_1, \ell_2} E[U(c_1, c_2, \ell_1, \ell_2)]$$

subject to the budget constraint

$$p_1 c_1 + \frac{p_2}{1+r} c_2 \leq w_1 \ell_1 + \frac{w_2}{1+r} \ell_2$$

where $p_1$ and $p_2$ are the prices of consumption goods in period 1 and 2, respectively, $w_1$ and $w_2$ are the nominal wage rates in the two periods, and r is the banks' lending and borrowing rate. In period 1, the agent does not yet know which price and wage level will prevail in period 2; $p_2$ and $w_2$ are thus to be considered as random variables when the decision of how much labor to supply in period 1 takes place. Solving this decision problem leads, according to Lucas and Rapping, to a labor supply function for period 1 of the form

$$\ell_1^* = \ell_1^*(w_1, w_2^e, p_1, p_2^e, r) \tag{7.2}$$

where $w_2^e$ and $p_2^e$ is the expected wage rate and price level, respectively, in period 2. This labor supply response is now transmitted through firms to the goods market, leading to an aggregate supply function (7.1) where $Y_t$ denotes the log of real output $y_t$, and $P_t$ denotes the log of the consumer price index $p_t$. The rational expectations approach to (7.1) is then to set $P_t^e \equiv E[P_t | P_{t-1}, \ldots]$ and see whether or not this implies that the natural rate hypothesis holds for the model in which (7.1) is utilized.[2]

In order to further investigate the roots of the rational expectations concept we go back to Muth's original (1961) paper. We recall that Muth's model was an ordinary model of market equilibrium, where the firms were subject to a one-period production lag, so that they had to decide at time t-1 (i.e. before they knew the price $p_t$) how much they would supply at time t:

$$\begin{aligned}
&\text{Demand:} && D_t = -\beta p_t \\
&\text{Supply:} && S_t = \gamma p_t^e + v_t \qquad (7.3)\\
&\text{Market equilibrium:} && D_t = S_t.
\end{aligned}$$

[2] Note however the ambiguity concerning whether $P_t^e$ should be interpreted according to

$$P_t^e \equiv E[\log p_t | \ldots]$$

or

$$P_t^e \equiv \log(E[p_t] | \ldots).$$

This has already been mentioned in the preceding chapter, and we shall not go more into that question here.

The parameters $\beta$ and $\gamma$ are positive constants, and $v_t$ is an error term corresponding to stochastic disturbances in the production process. We see thus that there is nothing strange with the model; the supply and demand curves have their usual slopes, and the producers who do not know $p_t$ in advance react on their forecast $p_t^e$ instead.

There is an important similarity between the supply functions (7.2) and (7.3): The agents do not know the future prices and wages, *but they act as if they knew them.* The only difference is that the unknown values of $w_2$, $p_2$, and $p_t$ are substituted by known, fixed numbers $w_2^e$, $p_2^e$, and $p_t^e$. Thus, the reaction functions are the same as they would be if there were no uncertainty involved in the problem, but next period's wages and prices were known for sure to be $w_2^e$, $p_2^e$, and $p_t^e$, respectively. In other words, it is assumed that there exist *certainty equivalents* for the uncertain variables. This is even explicitly mentioned by Muth, who writes that "The certainty-equivalence property follows from the linearity of the derivative of the appropriate quadratic profit or utility function".[3] The purpose of this chapter is to investigate the certainty equivalents within the framework of the Muth and Lucas-Rapping models. We will demonstrate that *for uncertain prices, the mathematical expectation is no certainty equivalent.*

## 7.1 THE CERTAINTY EQUIVALENCE PRINCIPLE

The economic literature contains two concepts referred to under the label "certainty equivalence". Firstly, we have the von Neumann-Morgenstern axiom that the individual's preference ordering over uncertain prospects is continuous; as a consequence of this axiom there exists for every lottery $\lambda$ a certain prospect x such that x is considered as equvalent to $\lambda$. Therefore x is often called the *certainty equivalent* of the uncertain prospect $\lambda$.

---

[3] Muth (1961, p. 317).

The second context in which the term certainty equivalence appears is related to the first one, but it has a much more specific interpretation. It is referred to as the *Certainty Equivalence Principle*, which states that the optimal solution of a particular class of decision problems under uncertainty is the same as the solution of the same problem under certainty, with the uncertain variables substituted by *known constants* which are equal to their conditional expectations. Due to the von Neumann-Morgenstern continuity axiom, there will always be a specific number which can replace the uncertain variable in the decision problem. This number will in general be a fairly complicated function of the parameters in the problem, the mean, variance and higher moments of the uncertain variable's probability distribution etc. The certainty equivalence principle states that for a particular class of decision problems, this number, the certainty equivalent, will be equal to the (conditional) expectation of the uncertain variable. Thus the probability distribution of the uncertain variable is irrelevant to the optimal solution (except for the mean of the distribution) and in particular the degree of uncertainty, as measured by e.g. the variance, does not matter.

This concept of certainty equivalence, which is due to Simon (1956) and Theil (1957),[4] is the one we will deal with in the present chapter. In its simplest form it can be stated as follows:

Suppose an agent has at his disposal a vector of m variables, to be called control variables or instruments, which are denoted $x \equiv (x_1\ x_2\ \dots\ x_m)^T$. By these variables he can affect n other variables, called state variables and denoted $y \equiv (y_1\ y_2\ \dots\ y_n)^T$. The vectors x and y are related according to the linear system

[4] More recent versions are given by Theil (1964, 1970). See also Malinvaud (1969).

$$y = Ax + \varepsilon \tag{7.4}$$

where $\varepsilon$ is an n-dimensional vector of stochastic disturbances $\varepsilon \equiv (\varepsilon_1\ \varepsilon_2\ \ldots\ \varepsilon_n)^T$ such that $\varepsilon$ is independent[5] of x. Suppose further that the agent wants to choose a control vector $(x_1\ \ x_2\ \ldots\ x_m)$ so as to maximize the expected value of a quadratic objective function

$$W \equiv E[y^TQy + x^TRx] \tag{7.5}$$

where Q and R are matrices of order n×n and m×m, respectively. Substituting (7.4) into (7.5) and taking the derivative with respect to x equal to zero yields the optimal instrument

$$x^* = E[-(A^TQ^TA + R^T)^{-1}A^TQ^T\varepsilon] = -(A^TQ^TA + R^T)^{-1}A^TQ^TE[\varepsilon].$$

We see that the uncertain variable $\varepsilon$ appears in $x^*$ only as a fixed number, the expectation $E[\varepsilon]$. In other words, the decision problem under *uncertainty* (7.4)-(7.5) will have the same solution as a similar problem under *certainty*, where y is given by the deterministic equation

$$y = Ax + E[\varepsilon].$$

This property, the certainty equivalence principle, thus holds for a simple, atemporal decision problem. However, it can also be generalized to hold, in a modified form, in a multi-period context. The generalization is fairly straightforward; for a $\bar{t}$-period decision problem we redefine the control variables as the $\bar{t}m$-dimensional vector

$$x \equiv (x_1\ x_2\ \ldots\ x_t\ \ldots\ x_{\bar{t}})^T$$

---

[5] For a discussion of what this independence assumption means in Theil's dynamic formulation of the principle, see Duchan (1974).

where

$$x_t \equiv (x_{1t}\ x_{2t}\ \cdots\ x_{mt})^T.$$

Similarly the state variables are redefined as an $\bar{t}n$-dimensional vector

$$y \equiv (y_1\ y_2\ \cdots\ y_t\ \cdots\ y_{\bar{t}})^T$$

where

$$y_t \equiv (y_{1t}\ y_{2t}\ \cdots\ y_{nt})^T.$$

By correspondingly extending the dimensionality of the A, Q and R matrices in the system (7.4) and the objective function (7.5), we obtain an intertemporal decision problem which displays what Theil calls *first period* certainty equivalence. This means that the first m elements of the optimal instrument $x^*$, i.e. $(x^*_{11}\ x^*_{21}\ \cdots\ x^*_{m1})^T$ are the same as they would be for a deterministic problem with the $\bar{t}n$-dimensional disturbance vector $\varepsilon$ replaced by its expectation conditional on all information available at time t=1.

Another way to approach the concept of certainty equivalence in a dynamic context is by means of the dynamic programming algorithm.[6] To apply dynamic programming methods, the system constraint (7.4) is often formulated as a vector-valued difference equation,

$$y_{t+1} = B_t\, y_t + C_t\, x_t + \varepsilon_t \qquad t = 1,\ 2,\ \ldots,\ \bar{t}-1$$

where the initial state y is given, and where the vector of stochastic disturbances $\varepsilon_t$ is independent of $x_t$. Assuming that

6 The two approaches are in fact equivalent, as is demonstrated by Norman (1974). For an account of stochastic dynamic programming, see e.g. Chow (1975) and Bertsekas (1976).

the objective function is quadratic and additive over time, this formulation leads to the optimal solution $x_t^*$ which appears not as a *fixed vector*, but as a *strategy*, a linear feedback function

$$x_t^* = L_t y_t + g_t. \qquad t = 1, 2, \ldots, t-1 \tag{7.6}$$

The matrices $L_t$ and the vectors $g_t$ are found by solving a fairly complicated difference equation. Now this expression clearly shows the multi-period nature of our solution; the values of $x_t^*$ cannot be set immediately in period 1 (except for t=1), for it depends on the realizations of $\varepsilon_{t-1}$, $\varepsilon_{t-2}$ etc., which have not yet been observed. But the optimal $x_t$ can be specified as a function of the $y_t$ value to appear at t, and is needed in the backward-optimization procedure of the dynamic programming algorithm to determine the optimal policies for earlier periods. When computing the matrices $L_t$ and the vectors $g_t$ it turns out that the expression for $L_t$ does not contain any disturbance term $\varepsilon_t$, and that $g_t$ is dependent only on the expectation $E[\varepsilon_t]$. We thus have a multi-period certainty equivalence in the sense that the optimal feedback rule (7.6) for the stochastic problem is identical to the solution of the corresponding deterministic problem with the uncertain variable $\varepsilon_t$ substituted by the known constant $E[\varepsilon_t]$.

The quadratic optimization problem under uncertainty can thus be generalized in a number of ways; it can be static or dynamic, and when dynamic, it can allow for transition matrices A and B, and objective weights Q and R, that are changing over time. Also, it can encompass the case when the number of objectives or the number of states is changing over time, i.e. when the vectors x and y are of dimensions $m_t$ and $n_t$, respectively. For all these generalizations the certainty equivalence holds in one form or another.

There are however some essential features in the problem (7.4)-(7.5) (which is general enough to encompass all the other generalizations) that constitute sufficient[7] conditions for the certainty equivalence principle to hold. These conditions are

i) A quadratic objective function

ii) Linear constraints

iii) Additive disturbances that are unaffected by the instruments.

These three requirements do obviously not encompass all problems of decisions under uncertainty. An important class of problems is constituted by cases where the coefficients in the constraints are uncertain, i.e. where *the matrix A in equation (7.4) is a random variable.*[8] For these cases the certainty equivalence principle does not hold in general, i.e. one cannot generally substitute the random matrix A by its expectation E[A] when computing the optimal control x*. This is evident from the solution of the problem (7.4)-(7.5); we recall that the optimal instrument was given by

$$x^* = E[-(A^TQ^TA + R^T)^{-1}A^TQ^T\varepsilon]$$

which, because of the fact that A appears in the solution in a non-linear way, is *not* necessarily equal to $-(E[A^T]Q^TE[A] + R^T)^{-1}E[A^T]Q^TE[\varepsilon]$. The random variable A can thus not be substituted by its mean, and the certainty equivalence principle does therefore not hold in the case with uncertain parameters.

---

[7] They are not necessary, since by sheer chance the principle might hold for some other problem which happens to have the appropriate parameter values.

[8] Optimal control of systems with unknown parameters is a fast-growing topic in the control literature; see e.g. Bertsekas (1976, p. 80-81) and Chow (1976).

Since prices in economic decision problems often appear as parameters in the constraints, i.e. as A above, price uncertainty is a special kind of uncertainty to which the certainty equivalence principle is not applicable. In the next section we will reconstruct the decision problems that lie behind models like the ones of Muth (1961) and Lucas-Rapping (1969), and we will see that the certainty equivalence principle is not applicable to these models.

## 7.2 MAXIMIZING BEHAVIOR UNDER PRICE UNCERTAINTY

Of the three standard assumptions usually made to prove the certainty equivalence principle, the first one, i.e. that the objective function be quadratic, is in some respects "permissible" in economic models. Of course we know that it constitutes a simplification which sometimes leads to strange results.[9] However, we will not question the first of the three assumptions on page 146 above, but concentrate on the other two.

Let us start with the simplest of all cases, a model with one produced commodity, one factor of production and a price-taking firm which has to decide today how much to produce to sell at tommorrow's market. The commodity is produced according to a production function $y = \phi(\ell)$, and the firm maximizes the expected utility of profit $\Pi$:

$$\max_{\ell} E[U(\Pi)] \equiv E[a\Pi^2 + b\Pi + c] \tag{7.7}$$

subject to

$$\Pi = p\phi(\ell) - w\ell \tag{7.8}$$

---

[9] See Borch (1969) and Feldstein (1969). Cf. also Tobin (1969).

where p is the price of the commodity, which is unknown when the production decision has to be taken, and w is the (known) price of the input. For the utility function U to be concave, we have that $a < 0$. For marginal utility $\frac{dU}{d\Pi}$ to be increasing, we have $b > 0$. We could also generalize the problem further by an additive stochastic term $\varepsilon$ in the constraint (7.8), corresponding to unforeseen disturbances in the production process, but no essential contributions to our main point are gained from such a complication.

This decision problem, although having an objective function which is quadratic in the state variable $\Pi$, does obviously not have a linear constraint in the instrument $\ell$. Thus the certainty equivalence principle is not immediately applicable to it. We have to simplify things a little further by assuming that $\phi(.)$ is a linear function $\phi(\ell) \equiv k \cdot \ell$. With no loss of generality we can set $k = 1$. The constraint (7.8) then reads

$$\Pi = (p-w)\ell. \tag{7.8'}$$

The economic interpretation of such a constraint is that the firm buys a quantity $\ell$ of a good today at a certain price w, stores it without cost overnight and sells it at an uncertain price p tomorrow. With these rather strong assumptions we have a decision problem with a quadratic objective function and a linear constraint. However, the stochastic disturbance does not appear additively, but in the system coefficient (p-w), corresponding to the system matrix A in the preceding section. Another interpretation would be to say that p consists of two parts, $\bar{p}$ which is certain and $\tilde{p}$ which is uncertain with mean zero. (7.8') could then be written

$$\Pi = (\bar{p}-w)\ell + \varepsilon$$

where $\varepsilon \equiv \tilde{p} \cdot \ell$. In this case the uncertainty is removed from the transition matrix A, and is added to the constraint in the usual way. But then $\varepsilon$ is *not independent of the instrument* $\ell$. Thus the third assumption is still violated.

In order to see the effects of an incorrect application of the certainty equivalence principle to the problem (7.7)-(7.8') we substitute the latter into the former, take the derivative with respect to $\ell$ equal to zero, and obtain the optimal instrument

$$\ell^* = \frac{b(w-E[p])}{2aE[p^2] + 2aw^2 - 4awE[p]} \equiv$$

$$\equiv \frac{b(w-E[p])}{2a\,Var[p] + 2a(w-E[p])^2} \tag{7.9}$$

which can be regarded as the firm's supply function of the good in question. For comparison, we assume for a while that the problem has no price uncertainty involved, i.e. p is a given constant, say $\alpha$. For this *hypothetical certainty case* the optimal instrument would be

$$\ell' = \frac{b}{2a(w-\alpha)} \tag{7.10}$$

For a strictly concave utility function we see that $\ell' > 0$ whenever $\alpha > w$.

Due to the von Neumann continuity axiom such a certainty equivalent exists. We can find its value by setting $\ell^*$, as given by (7.9), equal to $\ell'$, given by (7.10), and solve for $\alpha$. This yields

$$\alpha = \frac{Var[p]}{E[p]-w} + E[p].$$

Obviously the "correct" certainty equivalent, $\alpha$, is not equal to E[p] unless Var[p] = 0. But if (E[p] - w) increases, the difference between $\alpha$ and E[p] will decrease. The intuitive reason for this is simple: if the expected profit (E[p] - w) is very large, it does not matter much if the firm employs a fine tuning, utilizing the correct certainty equivalent $\alpha$, or employs rough rules-of-thumb, utilizing E[p] as a certainty equivalent.

If one makes the false analogy between the present decision problem and a problem to which the certainty equivalence principle really applies, one would substitute $\alpha$ in (7.10) by E[p] and then believe that the instrument

$$\hat{\ell} = \frac{b}{2a(w-E[p])}$$

thereby provides the optimal solution to the case with price uncertainty. We see then that $\hat{\ell} > \ell^*$ and that the difference $\hat{\ell} - \ell^*$ will increase with Var[p].

The above model provides a simple explanation of why the expected price cannot be a certainty equivalent. Its structure is perhaps a little bit artificial, but one should remember that more realistic assumptions (for example a concave production function $\phi(\ell)$) removes it even further from the conditions i)-iii) on page 146 above. Anyway, the model is perhaps similar to the one Muth had in mind when he formulated the firm's supply function in his 1961 paper; but since it is completely atemporal it is obviously not the kind of model that could be thought of as underlying the Lucas-Rapping (1969) model. For that sake we have to construct a two-period decision problem where an agent in period one does not yet know which price will prevail in period two. Contrary to the model (7.7)-(7.8'), however, he can decide to act *either* in period one *or* in period two; this model thereby has a dynamic framework which was lacking in the above model.[10]

---

[10] This is the kind of model Santomero and Seater (1978, p. 519 f.) refer to as "models in which the individual speculates over time".

Lucas' and Rapping's model requires a fairly complicated decision problem, however, with consumption at time 1, consumption at time 2, labor supply at time 1, labor supply at time 2, and savings between the two periods as decision variables. In order to demonstrate the main points, but still obtain manageable results, we form a simpler model where the agent has a resource to dispose of during the two periods at a price $p_1$ and $p_2$, respectively. The agent (which could be a firm owning an oil well, or a person who sells his labor) maximizes the expected utility of consumption during the two periods

$$\max_{x_1,x_2} E[U(c_1,c_2)] \equiv E\left[(c_1\ c_2)\begin{pmatrix}a_{11} & 0\\ 0 & a_{22}\end{pmatrix}\begin{pmatrix}c_1\\ c_2\end{pmatrix} + (c_1\ c_2)\begin{pmatrix}b_1\\ b_2\end{pmatrix}\right]. \tag{7.11}$$

The maximization is made subject to the linear constraints

$$\begin{aligned} c_1 &= p_1x_1 \\ c_2 &= p_2x_2 \\ x_1 + x_2 &= S \\ x_1 &\geq 0 \\ x_2 &\geq 0 \end{aligned} \tag{7.12}$$

where $x_1$ and $x_2$ are the quantities sold during the two periods and S is the initial stock of the resource.

This formulation implies several specific features of the world. Firstly we see that no money can be saved from period 1 to period 2; this is obviously an unrealistic assumption which is made solely to simplify the solution. Secondly we see that the price of consumer goods is certain, constant, and equal to unity for both periods. This means that we cannot analyze monetary phenomena like inflation, or like uncertainty of inflation. To illustrate our main point such limitations are inessential, but in order to lay microfoundations for a more complete macro model, complications like those must be taken into consideration.

Now solving the decision problem (7.11)-(7.12) yields two optimal instruments[11]

$$x_1^* = \frac{2Sa_{22}E[p_2^2] - b_1p_1 + b_2E[p_2]}{2a_{11}p_1^2 + 2a_{22}E[p_2^2]} \equiv$$

$$\equiv \frac{2Sa_{22}(E[p_2])^2 + 2Sa_{22}Var[p_2] - b_1p_1 + b_2E[p_2]}{2a_{11}p_1^2 + 2a_{22}(E[p_2])^2 + 2a_{11}Var[p_2]} \qquad (7.13)$$

$$x_2^* = S - x_1^*.$$

To trace the effects on x* of a change in prices, we first have to assume that the optimal consumption actually lies on the upward-sloping part of the utility parabola, i.e. that

$$\frac{\partial E[U(c_1^*, c_2^*)]}{\partial c_1} \equiv 2a_{11}p_1x_1^* + b_1 > 0. \qquad (7.14)$$

Now the sign of the partial derivative

$$\frac{\partial x_1^*}{\partial p_1} = \frac{2a_{11}b_1p_1^2 - 2a_{22}b_1E[p_2^2] - 8a_{11}a_{22}Sp_1E[p_2^2] - 4a_{11}b_2p_1E[p_2]}{(2a_{11}p_1^2 + 2a_{22}E[p_2^2])^2}$$

can be determined by taking (7.14), with $x_1^*$ given by (7.13), into account; it then turns out that

$$\frac{\partial x_1^*}{\partial p_1} > 0 \qquad \frac{\partial x_2^*}{\partial p_1} < 0$$

as was to be expected. Similarly we can show that

---

11 For simplicity we assume that the parameters are such that an interior solution obtains.

$$\frac{\partial x_1^*}{\partial E[p_2]} < 0 \qquad \frac{\partial x_2^*}{\partial E[p_2]} > 0.$$

A slightly more complicated task is to analyze the effects of an increased risk, i.e. to determine the sign of

$$\frac{\partial x_1^*}{\partial Var[p_2]} = \frac{2Sa_{22}D - 2a_{11}N}{D^2} \tag{7.15}$$

where N and D are the numerator and denominator, respectively, of $x_1^*$ given by (7.13). $D^2$ is of course positive. $a_{22}D$ is negative and so is $a_{11}N$; the sign of the difference $Sa_{22}D - a_{11}N$ is thus undetermined. However, assuming an interior solution we have $0 < x_1^* < S$, i.e.

$$0 < \frac{N}{D} < S$$

which gives $N > SD$. Multiplying both members by a negative number, say 2a, we obtain

$$2SaD - 2aN > 0.$$

Thus, if $a_{11} = a_{22} = a$, the expression (7.15) must be positive. It is also positive if $a_{11}$ is greater than $a_{22}$ in absolute value, which is the usual case in intertemporal models. In these the utility function is often assumed to be the same in both periods, except for the time preference which enters as a scale factor. The utility function in (7.11) is then specified as

$$\begin{aligned} E[U(c_1,c_2)] &\equiv E[U(c_1) + \frac{1}{1+\rho} U(c_2)] \\ &\equiv E[ac_1^2 + b + \frac{1}{1+\rho} ac_2^2 + \frac{1}{1+\rho} b]. \end{aligned}$$

For this case we then have $a_{11} \equiv a$ and $a_{22} = a/(1+p)$. Assuming $|a_{11}| \geq |a_{22}|$ is thus in conformity with economic praxis; we then have

$$\frac{\partial x_1^*}{\partial \mathrm{Var}[p_2]} > 0 \qquad \frac{\partial x_2^*}{\partial \mathrm{Var}[p_2]} < 0. \tag{7.16}$$

This is also what should be intuitively expected:[12] if the risk in tomorrow's uncertain market increases,[13] a risk-averse agent will rely relatively more on today's sure market, thereby increasing $x_1^*$ at the expense of $x_2^*$.

After having analyzed the correct solution, we turn our attention to the incorrect application of the certainty equivalence principle to the decision problem. If there were no price uncertainty, the optimal solution would be

$$\begin{aligned} x_1' &= \frac{2Sa_{22}p_2^2 - b_1p_1 + b_2p_2}{2a_{11}p_1^2 + 2a_{22}p_2^2} \\ x_2' &= S - x_1' \end{aligned} \tag{7.17}$$

Drawing the false analogy with a problem for which the certainty equivalence principle holds, the problem with uncertainty in $p_2$ is "solved" by replacing $p_2$ in (7.17) by $E[p_2]$. We thus have

$$\begin{aligned} \hat{x}_1 &= \frac{2Sa_{22}(E[p_2])^2 - b_1p_1 + b_2E[p_2]}{2a_{11}p_1^2 + 2a_{22}(E[p_2])^2} \\ \hat{x}_2 &= S - \hat{x}_1. \end{aligned}$$

Because of the inequalities (7.16), we see that

$$x_1^* > \hat{x}_1 \quad \text{and} \quad x_2^* < \hat{x}_2,$$

and that the differences $(x_1^* - \hat{x}_1)$ and $(\hat{x}_2 - x_2^*)$ tend to increase as $\mathrm{Var}[p_2]$ increases. Thus the incorrect application of the certainty equivalence principle yields a policy which is *too conservative* in the sense of saving too much of the resource for period 2.

---

[12] Cf. the analogous result for a slightly different model in Hoel (1977, p. 276).

[13] In our simple model we can identify the concept of "increasing risk" with an increase in $\mathrm{Var}[p_2]$. In general, however, these two concepts need not be equivalent. Cf. Rotschild and Stiglitz (1970, 1971).

## 7.3 CONCLUDING COMMENTS

We started this chapter by noticing that the supply equation $Y_t = a + b(P_t - P_t^e)$ had the "wrong" curvature in the sense that $E[y_t]$ was an increasing function of $Var[p_t]$. We have seen that it is incorrect in some way to include only a price expectations term in the function, thereby disregarding higher moments. This could be one explanation to why a curve derived from an incorrect application of the certainty equivalence principle has, when fitted to observed data, turned out to have that particular curvature.

One way to derive the supply function (7.1) stems from Lucas' and Rapping's (1969) two-period model of labor supply with uncertain wages and prices in the second period. We should not stretch the analogy between this model and our two-period model of optimal resource extraction too far, but just point out the possibility that the properties of (7.1) might have some similarities to the properties of $\hat{x}_1$, the supply function which follows from an incorrect application of the certainty equivalence principle. If the agents' behavior is given by a function $x_1^*(p_1, E[p_2], Var[p_2])$ and we try to estimate the observed time series $\{..., x_{1,t-1}^*, x_{1,t}^*, x_{1,t+1}^*, ...\}$ as a function of the time series $\{p_{1t}\}$ and $\{E[p_{2t}|p_{2t-1}]\}$ only, we are bound to get a functional form $\hat{x}_1(p_1, E[p_2])$ which implies that $E[\hat{x}_{1,t}]$ behaves strangely when $Var[p_{2,t}]$ is changed. Since the Lucas-Rapping model is so much richer and more complicated than our model, however, it defies the explicit solution which could give a definite answer to these questions. We should therefore not pursue this line of thought any further, but just point at the possibility that similar properties might exist.

Another way of obtaining a relation like the supply equation (7.1) is to develop a model like the one of Laidler (1978). There the firm speculates over markets, deciding how much to sell on market i by comparing the price $p_i$ to its conception of the price index taken over all markets $p^e$. This looks somewhat similar to our model of resource extraction; instead of distributing goods over time (where the price at faraway dates is uncertain), the firm distributes them over markets (where the price in faraway places is uncertain). Similar relations might hold concerning the partial derivatives, and this could be part of the explanation why the curvature of the "incorrect" supply function is the way it is. Here one should however be even more careful with the analogies, since Laidler's model is not cast in any time series context, like the Lucas-Rapping model and the resource extraction model are.

There is one obvious conclusion to be drawn from the above, namely that one should be suspicious of models involving price expectations terms. Since the certainty equivalent of an uncertain price is not equal to its mathematical expectation, functions of the type $y_t = a + bp_t^e$ (i.e. Muth's supply function) are not based on solid microfoundations. The question is of course whether or not this matters; in economic model building we have to make simplifying assumptions, and the application of the certainty equivalence principle to a case with price uncertainty is perhaps a permissible simplification - even if it can be criticized from a strictly theoretical point of view.

This might be true. For example, the expectations-augmented Phillips curve

$$\pi_t = f(u_t) + \alpha\pi_t^e$$

might be a good representation of reality in the sense that it conforms well to observed time series in inflation rates $\pi_t$ and unemployment rates $u_t$. One could therefore say that it is a permissible model on the aggregate level, even if it does not rely on firm foundations on the micro level. Similarly, a supply function of the type criticized in this chapter might be a permissible approximation to an aggregate reality, even if the individual firms' decision rules look quite different.

When dealing with *rational expectations models*, however, things are somewhat different. In the concept of "rational expectations" lies a strong emphasis on the assumption that agents are rational indeed, and that they are maximizers in the neoclassical tradition. Therefore the economic models of the rational expectations school should conform to the maximizing behavior of the agents; for a school which so strongly emphasizes rationality and maximizing behavior we must lay correspondingly strong claims on the econometric models utilized. To assume that the agents are rational in their formation of expectations, but not in their behavioral rules, seems highly questionable.[14] Therefore the incorrect application of the certainty equivalence principle might not be a permissible simplification in the particular case of rational expectations models.

---

[14] This point is also raised by Fair (1978).

# 8. SUMMARY AND CONCLUSIONS

It is difficult to model expectations. No representation of people's minds and hopes can be both *correct* and *operational*. We can therefore never give a "true" representation of inflationary expectations; what we can do is to make expectations models that are *reasonable* and operational. By reasonable we mean that our schemes have at least some trace of relevance and realism, and perhaps also that they have a well-established tradition in economic and econometric work. There are only two such expectations schemes, namely the adaptive expectations and the rational expectations.[1] Both are equally realistic (or unrealistic) but they are not completely taken out of the air - and they also have well-established traditions in economics. We have therefore studied these two schemes, to investigate whether the natural rate hypothesis is robust with respect to the modelling of expectations, i.e. whether the two expectations schemes have similar implications for the natural rate hypothesis.

We first found that the use of adaptive expectations within econometric studies might have been incorrect. If one

---

[1] It is of course not impossible that people's expectations are actually a combination of these two schemes. Gordon (1976, p. 204), for example, conceives of an expectations scheme which is a weighted average of rational and adaptive expectations:

$$\pi_t^e \equiv \lambda\, E[\pi_t | H_{t-1}] + (1-\lambda) \sum w_j \pi_{t-j}.$$

For the sake of simplicity, we have disregarded such permutations of the basic schemes.

wants to make a one-period forecast of a variable that follows a stationary stochastic process, then the optimal (in the sense of minimizing the mean square error) adaptive weight sum is less than unity. In earlier econometric work, the weights have always been constrained to sum to unity, and the conclusion is that this practice might have caused a bias in the estimates of other parameters within the models. In particular, since inflation actually seems to have followed a stationary time series over the years, estimates of the money illusion parameter $\alpha$ might have had a downward bias which might have led to an incorrect rejection of the natural rate of hypothesis.

There is one important qualification to be made, namely that one-period-ahead forecasts are perhaps not the most relevant ones for the Phillips curve model. As shown at the end of Chapter 4, multi-period forecasting is much more complicated than its one-period counterpart; the weight sum could be less than or greater than unity, and it could increase (as well as decrease) with the forecasting horizon.

In Chapter 5 we applied the concept of optimal adaptive expectations to a Phillips curve model to see whether it displayed natural rate properties. It turned out to do so, if the forecasting horizon was one period. With unemployment pegged at a rate different from the "natural" one, the inflationary process became more and more like a random walk with drift. Since a random walk with drift is an explosive process, this can be interpreted as a vertical long-run Phillips curve.

However, this natural rate result refers only to one-period forecasts. If the Phillips curve mirrors an economy dominated by multi-period contracts, things get more complicated. There are two different ways of representing such an economy, either by equation (5.13a), where the agents try to compensate themselves today for price changes expected to occur in the future, or (5.13b), where the agents entered long-term contracts $\theta$ periods ago, thereby determining today's inflation rate. These two formulations have different implications for the natural rate hypothesis (NRH). We saw by means of a simple numerical example that the latter displayed natural rate properties, while the former did not.

There is a lot of criticism to be raised against the concept of adaptive expectations, and therefore we also have to study the alternative, i.e. rational expectations. It is obvious that rational expectations, applied to the Phillips curve model of Chapter 5, implies the NRH. This is so because the conditional expectations operator $E[\cdot|\cdots]$ is an unbiased operator; taking the unconditional expectation of

$$\pi_t = f + E[\pi_{t+\theta}|\pi_{t-1}, \pi_{t-2}, \ldots]$$

or

$$\pi_t = f + E[\pi_t|\pi_{t-\theta}, \pi_{t-\theta-1}, \ldots]$$

gives $f = 0$ regardless of the forecasting horizon $\theta$. However, if we apply rational expectations to the *aggregate supply model*, which is the one generally used in rational expectations contexts, things get different. In Chapter 6 we noted, firstly, that there was an ambiguity of how the expectations term $P_t^e$ should be interpreted; in stochastic models the logarithm of the expectation and the expectation of the logarithm are quite different. We also noted that, contrary to what has sometimes been claimed, rational expectations are not equivalent to the NRH; neither of the two interpretations of $P_t^e$ implied the NRH. In particular, when we interpreted $P_t^e$ in the way which is most common in macro models, it turned out that the average of real output was an *increasing* function of the variance in the money supply rule. Since it is generally agreed that a highly erratic money supply rule is not desirable - it is supposed to have negative effects on planning and production within the firms - the conclusion is that the aggregate supply function has the "wrong" curvature.

In Chapter 7, finally, we tried to see what exactly was wrong with the aggregate supply function. When examining its micro-foundations, we found that the problem might come from an incorrect application of the certainty equivalence principle. We demonstrated that price uncertainty is a particular kind of uncertainty, such that the certainty equivalent of an uncertain price is not equal to its mathematical expectation. The conclusion from this is that rational expectations models, substituting an uncertain price by a variable $p_t^e$ defined by $p_t^e \equiv E[p_t|\ldots]$, do not rely on sound micro-foundations. This is perhaps one explanation of the strange properties of the aggregate supply curve.

There are two sets of conclusions to be drawn from the above results. One set refers to properties of widely used econometric models, and these conclusions are simple and straightforward. Our point of departure was that expectations are meaningful only if there is uncertainty involved, and that inflationary expectations should therefore be studied within the framework of stochastic macro models. As demonstrated in this book, the stochastic approach yields some insights that can not be obtained from deterministic models. We can conclude that if the time series to be predicted is stationary, the adaptive weight sum should be less than unity; that the aggregate supply function has the "wrong" curvature; that for uncertain prices the mathematical expectation is not a certainty equivalence, and the like.

The other set of conclusions refers to economic reality and the conduct of economic policy. What we would like to know is whether the natural rate hypothesis holds, or whether the authorities could exploit some kind of Phillips trade-

-off even in the long run. It is hardly surprising that such an important question has to be left without a simple answer. If the NRH had been robust with respect to the expectations mechanism (adaptive or rational) and with respect to the model formulation (Phillips curve or supply function) things would have been simpler. But now it turned out that our results pointed in different directions: Our result on one-period adaptive expectations supported the NRH, both concerning empirical estimates of the money illusion parameter $\alpha$ (Chapter 4) and concerning a theoretical analysis of inflationary processes (Chapter 5). On the other hand, our results on multi-period forecasting were inconclusive. Turning to the rational expectations, our results was clearly against the NRH, but also indicated that the aggregate supply function should not be trusted, since it is probably mis-specified.

The conclusion is therefore that the NRH is not particularly robust. Using alternative, equally reasonable, models yields different results, and decisive tests are not yet to be obtained.

# REFERENCES


Akhiezer, N.I.: *The Classical Moment Problem*, Oliver & Boyd, London 1965.

Almon, S.: "The Distributed Lag Between Capital Appropriations and Expenditures", *Econometrica*, 1965, Vol. 33, pp. 178-196.

Bailey, M.J.: "Prediction of an Autoregressive Variable Subject Both to Disturbances and to Errors of Observation", *Journal of the American Statistical Association*, March 1965, Vol. 60, No. 309, pp. 164-181.

____________: *National Income and the Price Level*, 2nd ed., McGraw-Hill, New York 1971.

Barro, R.J.: "Rational Expectations and the Role of Monetary Policy", *Journal of Monetary Economics*, January 1976, Vol. 2, No. 1, pp. 1-32.

Barro, R.J. and S. Fischer: "Recent Developments in Monetary Theory", *Journal of Monetary Economics*, April 1976, Vol. 2, No. 2, pp. 133-167.

Barro, R.J. and H. Grossman: *Money, Employment and Inflation*, Cambridge University Press, Cambridge 1976.

Bertsekas, D.P.: *Dynamic Programming and Stochastic Control*, Academic Press, New York 1976.

Black, F.: "Uniqueness of the Price Level in Monetary Growth Models with Rational Expectations", *Journal of Economic Theory*, January 1974, Vol. 7, No. 1, pp. 53-65.

Borch, K.: "A Note on Uncertainty and Indifference Curves", *Review of Economic Studies*, January 1969, Vol. XXXVI, No. 1, pp. 1-4.

Box, G.E.P. and G.M. Jenkins: *Time Series Analysis*, Holden--Day, San Francisco, 1970. 2nd ed. 1976.

Brock, W.A.: "Money and Growth: The Case of Long Run Perfect Foresight", *International Economic Review*, October 1974, Vol. 15, No. 3, pp. 750-777.

Burmeister, E. and S.J. Turnovsky: "The Specification of Adaptive Expectations in Continuous Time Dynamic Economic Models", *Econometrica*, September 1976, Vol. 44, No. 5, pp. 879-905.

Cagan, P.: "The Monetary Dynamics of Hyperinflation", in M. Friedman et al. (1956).

Chow, G.C.: *Analysis and Control of Dynamic Economic Systems*, John Wiley & Sons, New York 1975.

__________: "The Control of Nonlinear Econometric Systems with Unknown Parameters", *Econometrica*, July 1976, Vol. 44, No. 4, pp. 685-695.

Clower, R.W.: "The Keynesian Counterrevolution: A Theoretical Appraisal", in F. Hahn, F. Brechling et al. (1965).

Cox, D.R. and H.D. Miller: *The Theory of Stochastic Processes*, Methuen & Co., London 1965.

Dhrymes, P.J.: *Distributed Lags: Problems of Estimation and Formulation*, Holden-Day, San Francisco 1971.

Duchan, A.L.: "A Clarification and a New Proof of the Certainty Equivalence Theorem", *International Economic Review*, February 1974, Vol. 15, No. 1, pp. 216-224.

Fair, R.C.: "A Criticism of a Class of Macroeconomic Models with Rational Expectations", *Journal of Money, Credit and Banking*, November 1978, Vol. 10, No. 4, pp. 411-417.

Feldstein, M.S.: "Mean-Variance Analysis in the Theory of Liquidity Preference and Portfolio Selection", *Review of Economic Studies*, January 1969, Vol. XXXVI, No. 1, pp. 5-12.

Feller, W.: *An Introduction to Probability Theory and Its Applications, Volume II*, Wiley, New York 1971.

Fischer, S.: "Long-Term Contracts, Rational Expectations, and the Optimal Money Supply Rule", *Journal of Political Economy*, February 1977, Vol. 85, No. 1, pp. 191-205.

Fisher, I.: "Our Unstable Dollar and the So-called Business Cycle", *Journal of the American Statistical Association*, June 1925, Vol. 20, pp. 179-202.

__________: "A Statistical Relation Between Unemployment and Price Changes", *International Labour Review*, June 1926, Vol. 13, pp. 785-792. Reprinted in *Journal of Political Economy*, March/April 1973, Vol. 81, pp. 496-502.

__________: *The Theory of Interest*, Macmillan, New York 1930.

Friedman, B.M.: "Stability and Rationality in Models of Hyperinflation", Harvard Institute of Economic Research, Discussion Paper 445, November 1975 (mimeo).

Friedman, B.M.: "Survey Evidence on the 'Rationality' of Interest Rate Expectations", Harvard Institute of Economic Research, Discussion Paper No. 635, July 1978 (mimeo).

______: "Optimal Expectations and the Extreme Information Assumptions of 'Rational Expectations' Macromodels", *Journal of Monetary Economics*, January 1979, Vol. 5, No. 1, pp. 23-41.

Friedman, M. (ed.): *Studies in the Quantity Theory of Money*, The University of Chicago Press, Chicago 1956.

Friedman, M.: *A Theory of the Consumption Function*, National Bureau of Economic Research, New York 1957.

______: *A Programme for Monetary Stability*, Fordham University Press, New York 1960.

______: " The Role of Monetary Policy", *American Economic Review*, 1968, Vol. 58, No. 1, pp. 1-17.

______, Nobel Lecture: "Inflation and Unemployment", *Journal of Political Economy*, June 1977, Vol. 85, No. 3, pp. 451-472.

Gordon, R.J.: "Recent Developments in the Theory of Inflation and Unemployment", *Journal of Monetary Economics*, 1976, Vol. 2, No. 2, pp. 185-219.

Granger, C.W.J.: "The Typical Spectral Shape of an Economic Variable", *Econometrica*, January 1966, Vol. 34, No. 1, pp. 150-161.

Granger, C.W.J. and M. Hatanaka: *Spectral Analysis of Economic Time Series*, Princeton University Press, Princeton 1964.

Granger, C.W.J. and P. Newbold: *Forecasting Economic Time Series*, Academic Press, New York 1977.

Griliches, Z.: "Distributed Lags: A Survey", *Econometrica*, January 1967, Vol. 35, No. 1, pp. 16-49.

Hahn, F. and F. Brechling (eds.): *The Theory of Interest Rates*, Macmillan, London 1965.

Hoel, M.: "Five Essays on the Extraction of an Exhaustible Resource Which Has a Non-Exhaustible Substitute", Memorandum from the Institute of Economics, University of Oslo, October 1977 (mimeo).

Holt, C.C.: "Job Search, Phillips' Wage Relation, and Union Influence: Theory and Evidence", in Phelps et al. (1970).

Jenkins, G.M. and D.G. Watts: *Spectral Analysis and Its Applications*, Holden-Day, San Francisco 1968.

Johansson, Ö.: *The Gross Domestic Product of Sweden and Its Composition 1861-1955*, Almqvist & Wiksell, Stockholm 1967.

Karlin, S. and H.M. Taylor: *A First Course in Stochastic Processes*, Academic Press, New York 1975.

Kendall, M.G. and A. Stuart: *The Advanced Theory of Statistics, Vol. 1*, Griffin & Co., London 1963.

Khan, M.S.: "The Variability of Expectations in Hyperinflations", *Journal of Political Economy*, August 1977, Vol. 85, No. 4, pp. 817-827.

Koyck, L.M.: *Distributed Lags and Investment Analysis*, North-Holland Publishing Co., Amsterdam 1954.

Laidler, D.E.W.: *Essays on Money and Inflation*, Manchester University Press, Manchester 1975.

______________: "Money and Money Income: An Essay on the 'Transmission Mechanism'", *Journal of Monetary Economics*, April 1978, Vol. 4, No. 2, pp. 151-191.

Lindbeck, A.: "Stabilization Policy in Open Economies With Endogeneous Politicians", *American Economic Review*, May 1976, Vol. 66, No. 2, pp. 1-19.

Lipsey, R.G.: "The Relation Between Unemployment and the Rate of Change of Money Wages in the United Kingdom, 1862-1957: A Further Analysis", *Economica*, February 1960, Vol. 27, No. 1, pp. 1-31.

Lucas, R.E.: "Expectations and the Neutrality of Money", *Journal of Economic Theory*, April 1972, Vol. 4, No. 2, pp. 103-124 (a).

__________: "Econometric Testing of the Natural Rate Hypothesis", in O. Eckstein (ed.): *The Econometrics of Price Determination*, Board of Governors of the Federal Reserve System, Washington 1972 (b).

__________: "Some International Evidence on Output-Inflation Tradeoffs", *American Economic Review*, June 1973, Vol. 63, No. 3, pp. 326-334.

__________: "Econometric Policy Evaluation: A Critique", in K. Brunner and A. Meltzer (eds.): *The Phillips Curve and Labor Markets*, Vol. 1 of the Carnegie-Rochester conference series on public policy, North-Holland, Amsterdam 1976 (a).

__________: "Some International Evidence on Output-Inflation Tradeoffs: Errata", *American Economic Review*, December 1976, Vol. 66, No. 5, p. 985 (b).

Lucas, R.E. and L.A. Rapping: "Real Wages, Employment, and Inflation", *Journal of Political Economy*, September/October 1969, Vol. 77, No. 5, pp. 721-751.

Luenberger, D.G.: *Optimization by Vector Space Methods*, Wiley, New York 1969.

Lundberg, E.: *Studies in the Theory of Economic Expansion*, King & Son, London 1937.

Malinvaud, E.: "First Order Certainty Equivalence", *Econometrica*, October 1969, Vol. 37, No. 4, pp. 706-718.

McNees, S.K.: "The 'Rationality' of Economic Forecasts", *American Economic Review*, May 1978, Vol. 68, No. 2, pp. 301-305.

Mincer, J. (ed.): *Economic Forecasts and Expectations*, National Bureau of Economic Research, New York 1969.

Mincer, J.: "Models of Adaptive Forecasting", in Mincer et al. (1969).

Muth, J.F.: "Optimal Properties of Exponentially Weighted Forecasts", *Journal of the American Statistical Association*, June 1960, Vol. 55, No. 290, pp. 299-306.

__________: "Rational Expectations and the Theory of Price Movements", *Econometrica*, July 1961, Vol. 29, No. 3, pp. 315-335.

Nelson, C.R.: *Applied Time Series Analysis for Managerial Forecasting*, Holden-Day, San Francisco 1973.

__________: "Rational Expectations and the Predictive Efficiency of Economic Models", *Journal of Business*, July 1975, Vol. 48, No. 3, pp. 331-343.

Nerlove, M.: *Distributed Lags and Demand Analysis*, Agricultural Handbook No. 141, U.S. Government Printing Office, Washington 1958.

__________: "Distributed Lags and Unobserved Components in Economic Time Series", in W. Fellner et al.: *Ten Economic Studies in the Tradition of Irving Fisher*, Wiley, New York 1967.

__________: "Lags in Economic Behavior", *Econometrica*, March 1972, Vol. 40, No. 2, pp. 221-251.

Nordhaus, W.D.: "The Political Business Cycle", *Review of Economic Studies*, April 1975, Vol. XLII, No. 2, pp. 169-190.

Norman, A.L.: "On the Relationship Between Linear Feedback Control and First Period Certainty Equivalence", *International Economic Review*, February 1974, Vol. 15, No. 1, pp. 209-215.

Persson, M.: "A Result on the Covariance Matrix of a Stationary AR Process", Economic Research Institute at the Stockholm School of Economics, Working Paper, May 1979 (mimeo).

Phelps, E.S.: "Phillips Curves, Expectations of Inflation, and Optimal Inflation Over Time", *Economica*, August 1967, Vol. 34 (NS), pp. 254-281.

Phelps, E.S., et al.: *Macroeconomic Foundations of Employment and Inflation Theory*, Norton, New York 1970.

Phelps, E.S.: *Inflation Policy and Unemployment Theory*, Norton, New York 1972.

Phelps, E.S. and J.B. Taylor: "Stabilizing Powers of Monetary Policy Under Rational Expectations", *Journal of Political Economy*, February 1977, Vol. 85, No. 1, pp. 163-190.

Phillips, A.W.: "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1862-1957", *Economica*, November 1958, Vol. 25, pp. 283--299.

Rothschild, M. and J.E. Stiglitz: "Increasing Risk: I. A Definition", *Journal of Economic Theory*, February 1970, Vol. 2, No. 1, pp. 225-243.

____________________: "Increasing Risk II: Its Economic Consequences", *Journal of Economic Theory*, February 1971, Vol. 3, No. 1, pp. 66-84.

Rutledge, J.: *A Monetarist Model of Inflationary Expectations*, Lexington Books, Lexington (Mass.) 1974.

Samuelson, P.A.: "Interactions Between the Multiplier Analysis and the Principle of Acceleration", *The Review of Economics and Statistics*, May 1939, Vol. XXI, No. 2, pp. 75-78.

Santomero, A.M. and J.J. Seater: "The Inflation-Unemployment Tradeoff: A Critique of the Literature", *Journal of Economic Literature*, June 1978, Vol. XVI, No. 2, pp. 499-544.

Sargent, Th.: "A Note on the 'Accelerationist' Controversy", *Journal of Money, Credit and Banking*, August 1971, Vol. 3, No. 3, pp. 721-725.

________: "'Rational' Expectations, the Real Rate of Interest, and the 'Natural' Rate of Unemployment", *Brookings Papers on Economic Activity*, No. 2, Washington, Brookings Institution 1973, pp. 429-480.

Sargent, Th. and N. Wallace: "Rational Expectations and the Dynamics of Hyperinflation", *International Economic Review*, June 1973, Vol. 14, No. 2, pp. 328-350.

________________: "'Rational' Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule", *Journal of Political Economy*, April 1975, Vol. 83, No. 2, pp. 241-254.

Seater, J.J.: "A Unified Model of Consumption, Labor Supply, and Job Search", *Journal of Economic Theory*, April 1977, Vol. 14, No. 2, pp. 349-372.

________: "Utility Maximization, Aggregate Labor Force Behavior, and the Phillips Curve", *Journal of Monetary Economics*, November 1978, Vol. 4, No. 4, pp. 687-713.

Shiller, R.J.: "Rational Expectations and the Dynamic Structure of Macroeconomic Models - A Critical Review", *Journal of Monetary Economics*, January 1978, Vol. 4, No. 1, pp. 1-44.

Sidrauski, M.: "Inflation and Economic Growth", *Journal of Political Economy*, December 1967, Vol. LXXV, No. 5, pp. 796-810.

Simon, H.A.: "Dynamic Programming Under Uncertainty With a Quadratic Criterion Function", *Econometrica*, January 1956, Vol. 24, No. 1, pp. 74-81.

Slutsky, E.: "The Summation of Random Causes as the Source of Cyclic Processes", *Econometrica*, 1937, Vol. 5, No. 2, pp. 105-146.

Solow, R.M.: *Price Expectations and the Behavior of the Price Level*, Manchester University Press, Manchester 1969.

Taylor, J.B.: "Monetary Policy During a Transition to Rational Expectations", *Journal of Political Economy*, October 1976, Vol. 83, No. 5, pp. 1009-1021.

Theil, H.: "A Note on Certainty Equivalence in Dynamic Planning", *Econometrica*, April 1957, Vol. 25, No. 2, pp. 346-349.

_________: *Optimal Decision Rules for Government and Industry*, North-Holland, Amsterdam 1964.

_________: *Economic Forecasts and Policy*, 2nd edition, North-Holland, Amsterdam 1970.

Tinbergen, J.: "Suggestions on Quantitative Business Cycle Theory", *Econometrica*, July 1935, Vol. III, No. 3, pp. 241-308.

Tobin, J.: "Comment on Borch and Feldstein", *Review of Economic Studies*, January 1969, Vol. XXXVI, No. 1, pp. 13-14.

Turnovsky, S.J.: "The Expectations Hypothesis and the Aggregate Wage Equation: Some Empirical Evidence for Canada", *Economica*, 1972, Vol. XXXIX (NS), No. 153, pp. 1-17.

Wold, H.: *A Study in the Analysis of Stationary Time Series*, 2nd edition, Almqvist & Wiksell, Uppsala 1954.

Yaglom, A.M.: *An Introduction to the Theory of Stationary Random Functions*, Prentice-Hall, Englewood Cliffs, N.J. 1962.

Åmark, K.: "En svensk prishistorisk studie", *Ekonomisk Tidskrift*, 1921, Vol. 23, No. XII, pp. 147-170.

LIST OF REPORTS PUBLISHED SINCE 1976 BY THE ECONOMIC RESEARCH INSTITUTE AT THE STOCKHOLM SCHOOL OF ECONOMICS


1976

BRODIN, B., Produktutvecklingsprocesser - En studie ur marknadsföringssynvinkel av produktutveckling i svenska företag. Stockholm 1976.
Del I - Analys
Del II - Praktikfallsbeskrivningar

BRODIN, B., JÖNSSON, S., NYSTRÖM, H. & SANDKULL, B., Fyra uppsatser om produktutveckling. Stockholm 1976. EFI/MTC.

DOCHERTY, P., Forskarroller i ett aktionsforskningsprojekt. Stockholm 1976. EFI/PA-rådet.

FALK, T., Urban Sweden - Changes in the distribution of population - the 1960s in focus. Stockholm 1976.

HEDEBRO, G., Information och engagemang - Individuella och miljömässiga förutsättningar för deltagande i det lokala samhällslivet: exemplet skola. Stockholm 1976.

HEDEBRO, G. & NOWAK, K., Studiecirklar om U-landsfrågor - Undersökning av studiecirklar med internationell inriktning i Göteborgs kommun våren 1975. Stockholm 1976. Stencil.
Rapport nr 1: Beskrivning av deltagarna
Rapport nr 2: Deltagarna och cirkelarbetet
Rapport nr 3: Ett halvt år efter cirkelarbetet
Rapport nr 4: Sammanfattande analys och diskussion

HEDLUND, G., Det multinationella företaget, nationalstaten och fackföreningarna - En diskussion av utgångspunkter och metoder. Stockholm 1976.

JULANDER, C-R., Evaluation of Consumer Information. Stockholm 1976. Stencil.

JULANDER, C-R., Utvärdering av möbelfakta. Stockholm 1976. Stencil.

LINDH, L.G., Dagligvaruleverantörernas samaktiviteter. Stockholm 1976.

PETERSSOHN, E., Kreditgivning mellan företag - En mikroekonometrisk studie av företagens finansiella beteende. Stockholm 1976.

RYDÉN, I., Transportkostnader och regional utveckling - Modeller för analys av regionalpolitiskt stöd av godstransporter. Stockholm 1976.

STOLT, B., Dialog mellan medborgare och myndigheter - En utopi? Effekter av myndighetsinformation med återföringssyfte. Stockholm 1976. Stencil.

STOLT, B., Om samhällskommunikation: Deskriptiva, stipulativa och normativa utgångspunkter för informationsutbyte mellan medborgare och myndigheter samt ett klassificeringsschema för individens politiska resurser. Stockholm 1976. Stencil.
TELL, B., A Comparative Study of Some Multiple-Criteria Methods. Stockholm 1976.
ÅHRÉN, P., Economic Evaluation Methods in Community Planning. Stockholm 1976. Statens råd för byggnadsforskning.

1977

ANELL, B., FALK, T., FJAESTAD, B., JULANDER, C-R, KARLSSON, T., STJERNBERG, T., Detaljhandel i omvandling. Studiens föreställningsram, uppläggning och genomförande. En rapport från Projektet "Detaljhandel i omvandling". Stockholm 1977.
BERGMAN, L., Energy and Economic Growth in Sweden - An analysis of historical trends and present choices. Stockholm 1977. Stencil.
BERTMAR, L., MOLIN, G., Kapitaltillväxt, kapitalstruktur och räntabilitet. En analys av svenska industriföretag. Stockholm 1977.
DOCHERTY, P., MAGNUSSON, A., STYMNE, B., CALLBO, K. & HERBER, S., Hur man lyckas med systemutveckling. En analys av fem praktikfall. Stockholm 1977.
EKLUND, L. & SJÖSTRAND, S-E., Inlärningsorganisation - Fallet Kista, Stockholm. Stockholm 1977.
FJAESTAD, B. & HOLMLÖV, P.G., Marknaden för bredbandstjänster - Intresse för och planerad medverkan i Televerkets provnät för bildtelefon, snabbfaksimil och special-tv. Stockholm 1977. Stencil.
HAMMARKVIST, K-O., Köpprocessen för nya produkter på byggmarknaden. Stockholm 1977.
HAMMARKVIST, K-O., Köparbeteende i byggbranschen. Två studier av adoptionsprocessen för byggmaterial. Stockholm 1977. Stencil.
HEDBERG, P., JOHANSSON, L. & JUNDIN, S., Barn och marknadskommunikation - Undersökning av yngre barns produktönskningar samt påverkanskällor till dessa. Stockholm 1977. Stencil.
HULTCRANTZ, G., LINDHOFF, H., VALDELIN, J., Chinese Economic Planning: Characteristics, Objectives and Methods - An Introduction. China's Developmental Strategy: 7. Stockholm 1977.
HÖGLUND, M., Sortimentsförändringar i detaljhandeln - En studie av svensk sällanköpsvaruhandel. Stockholm 1977. EFI/MTC.
JULANDER, C-R. & LÖÖF, P-O., Utvärdering av möbelfakta i konsumentledet. Stockholm 1977.
JULANDER, C-R. & LÖÖF, P-O., Utvärdering av möbelfakta i producent- och detaljistledet. Stockholm 1977.
JULANDER, C-R., LÖÖF, P-O. & LINDQVIST, A., Beslutsunderlag för utformning av varuinformation - Intervjuer med konsumenter och analys av reklamationer i möbelbranschen. Stockholm 1977.
JULANDER, C-R., LINDQVIST, A., FJAESTAD, B., Utveckling av beteendevetenskapliga indikatorer på sparande - En litteraturöversikt samt en intervjuundersökning av 50 stockholmshushåll. Rapport nr 1 från projektet "Sparande och sparbeteende". Stockholm 1977.
JULANDER, C-R., Effekter av jämförande varuprovningar av diskmaskiner i producentledet. Stockholm 1977.

KARLSSON, H., Köpcentret Linden - Genomförande och utformning. En rapport från projektet "Detaljhandel i omvandling". Stockholm 1977.
MOSSBERG, T., Utveckling av nyckeltal. Stockholm 1977.
OLVE, N-G., Multiobjective Budgetary Planning - Models for interactive planning in decentralized organizations. Stockholm 1977.
STJERNBERG, T., Organizational Change and Quality of Life - Individual and Organizational perspectives on democratization of work in an insurance company. Stockholm 1977.
ÖSTMAN, L., Styrning med redovisningsmått - En studie av resultat- och räntabilitetsmåtts användning i divisionaliserade företag. Stockholm 1977.

1978

ASPLING, A. & LINDESTAD, L., Företagsutbildning. Reflektioner kring tre praktikfall. Stockholm 1978.
EKMAN, Elon V., Some Dynamic Economic Models of the Firm. A microeconomic analysis with emphasis on firms that maximize other goals than profit alone. Stockholm 1978.
FJAESTAD, B., JULANDER, C-R., Norrköpingshandelns syn på framtiden. En studie av köpmännens förväntningar rörande Linden-etableringens effekter på konsumenternas beteende. Rapport nr 6 från forskningsprogrammet "Detaljhandel i omvandling". Stockholm 1978. Stencil.
FJAESTAD, B., JULANDER, C-R., ANELL, B., Konsumenter i Norrköpings köpcentra. En studie av reseavstånd, färdmedel, ärenden och missnöje. Rapport nr 4 från forskningsprogrammet "Detaljhandel i omvandling". Stockholm 1978.
HEMMING, T., Multiobjective Decision Making under Certainty. Stockholm 1978.
HOLMLÖV, P.G., Lokalpressen och kommunalpolitiken. Stockholm 1978.
JULANDER, C-R., EDSTRÖM, E., Beslutsunderlag för utformning av varuinformation om bostäder. En intervjuundersökning med hyresgäster i Husby/Akalla. Stockholm 1978. Stencil.
LINDQVIST, A., JULANDER, C-R. & FJAESTAD, B., Utveckling av beteendevetenskapliga indikatorer på sparande. Rapport nr 2 från projektet "Sparande och sparbeteende". Stockholm 1978.
SCHWARZ, B. & SVENSSON, J-E., Transporter och transportforskning. Ett framtidsperspektiv. Stockholm 1978.
STJERNBERG, T., Detaljhandelsanställda i Linden. En studie av de anställdas inflytande och upplevelser i samband med etableringen av ett köpcentrum. Rapport nr 3 från forskningsprogrammet "Detaljhandel i omvandling". Stockholm 1978. Stencil.
TELL, B., Investeringskalkylering i praktiken. Stockholm 1978.
Människan i organisationen. Red.: Sjöstrand, S-E., Stymne, B. & Forsblad, P., Stockholm 1978.
Telefonen som brevlåda. En studie av 100 svenska användare av telefaksimil. Red.: Jundin, S. & Lindqvist, A., Stockholm 1978.

1979

AHLMARK, D. & BRODIN, B., Bokbranschen i framtiden - En analys av möjliga utvecklingslinjer fram till år 2000. Stockholm 1979.

AHLMARK, D. & BRODIN, B., Statligt litteraturstöd - En ekonomisk analys. Stockholm 1979.

AHLMARK, D. & LJUNGKVIST, M-O., Finansiell analys och styrning av bokförlag. Studier av utveckling och beteende under 1970-talet. Stockholm 1979.

ANELL, B., När butiken försvinner. Rapport nr 5 från forskningsprogrammet "Detaljhandel i omvandling". Stockholm 1979.

FALK, T., Detaljhandeln i Norrköping. Struktur och lokalisering 1977. Rapport nr 7 från forskningsprogrammet "Detaljhandel i omvandling". Stockholm 1979.

HOLMLÖV, P.G., FJAESTAD, B. & JULANDER, C-R., Form och funktion i marknadsföringen. Försäljning, produktutveckling och reklamargumentation på marknaderna för vitvaror och snickerier i kök 1961-1976. Stockholm 1979.

JULANDER, C-R. & FJAESTAD, B., Konsumenternas köpvanor i Norrköping och Söderköping. Rapport nr 8 från forskningsprogrammet "Detaljhandel i omvandling". Stockholm 1979.

MAGNUSSON, Å., PETERSSOHN, E. & SVENSSON, C., Sakförsäkring och inflation. En ekonomisk analys. Stockholm 1979.

ÖSTERBERG, H., Hierarkisk begreppsanalys. Ett hjälpmedel vid undersökning av komplexa forskningsproblem. Stockholm 1979.