Note 2.

10.10.2016 Christian Groth

Uncertainty, expectations, and asset price bubbles

This lecture note provides a framework for addressing themes where expectations in *uncertain* situations are important elements. Our previous models have not taken seriously the problem of uncertainty. Where agent's expectations about future variables were involved and these expectations were assumed to be model-consistent ("rational"), we only considered a special case: perfect foresight. Shocks were treated in a peculiar (almost self-contradictory) way: they might occur, but only as a complete surprise, a one-off event. Agents' expectations and actions never incorporated that new shocks could arrive.

We will now allow recurrent shocks to take place. The environment in which the economic agents act will be considered inherently uncertain. How can this be modeled and how can we solve the resultant models? Since it is easier to model uncertainty in discrete rather than continuous time, we examine uncertainty and expectations in a discrete time framework.

Our emphasis will be on the hypothesis that when facing uncertainty a dominating fraction of the economic agents form "rational expectations" in the sense of making probabilistic forecasts which coincide with the forecast calculated on the basis of the "relevant economic model". But we begin with simple mechanistic expectation formation hypotheses that have been used to describe day-to-day expectations of people who do not think much about the probabilistic properties of their economic environment.

1 Simple expectation formation hypotheses

One simple supposition is that expectations change gradually to correct past expectation errors. Let P_t denote the general price level in period t and $\pi_t \equiv (P_t - P_{t-1})/P_{t-1}$ the corresponding inflation rate. Further, let $\pi^e_{t-1,t}$ denote the "subjective expectation", formed in period $t-1$, of π_t, i.e., the inflation rate from period $t-1$ to period t. We may think of the "subjective expectation" as the expected value in a vaguely defined subjective conditional probability distribution.

The hypothesis of *adaptive expectations* (the AE hypothesis) says that the expectation is revised in proportion to the past expectation error,

$$\pi_{t-1,t}^e = \pi_{t-2,t-1}^e + \lambda(\pi_{t-1} - \pi_{t-2,t-1}^e), \; 0 < \lambda \leq 1, \tag{1}$$

where the parameter λ is called the adjustment speed. If $\lambda = 1$, the formula reduces to

$$\pi_{t-1,t}^e = \pi_{t-1}. \tag{2}$$

This limiting case is known as *static expectations* or *myopic expectations;* the subjective expectation is that the inflation rate will remain the same. As we shall see, *if* inflation follows a random walk, this subjective expectation is in fact the "rational expectation".

We may write (1) on the alternative form

$$\pi_{t-1,t}^e = \lambda\pi_{t-1} + (1-\lambda)\pi_{t-2,t-1}^e. \tag{3}$$

This says that the expected value concerning this period (period t) is a weighted average of the actual value for the last period and the expected value for the last period. By backward substitution we find

$$\begin{aligned}
\pi_{t-1,t}^e &= \lambda\pi_{t-1} + (1-\lambda)[\lambda\pi_{t-2} + (1-\lambda)\pi_{t-3,t-2}^e] \\
&= \lambda\pi_{t-1} + (1-\lambda)\lambda\pi_{t-2} + (1-\lambda)^2[\lambda\pi_{t-3} + (1-\lambda)\pi_{t-4,t-3}^e] \\
&= \lambda\sum_{i=1}^{n}(1-\lambda)^{i-1}\pi_{t-i} + (1-\lambda)^n\pi_{t-n-1,t-n}^e.
\end{aligned}$$

Since $(1-\lambda)^n \to 0$ for $n \to \infty$, we have (for $\pi_{t-n-1,t-n}^e$ bounded as $n \to \infty$),

$$\pi_{t-1,t}^e = \lambda\sum_{i=1}^{\infty}(1-\lambda)^{i-1}\pi_{t-i}. \tag{4}$$

Thus, according to the AE hypothesis with $0 < \lambda < 1$, the expected inflation rate is a weighted average of the historical inflation rates back in time. The weights are geometrically declining with increasing time distance from the current period. The weights sum to one (in that $\sum_{i=1}^{\infty}\lambda(1-\lambda)^{i-1} = \lambda(1-(1-\lambda))^{-1} = 1$).

The formula (4) can be generalized to the *general backward-looking expectations* formula,

$$\pi_{t-1,t}^e = \sum_{i=1}^{\infty} w_i\pi_{t-1-i}, \qquad \text{where } \sum_{i=1}^{\infty} w_i = 1. \tag{5}$$

If the weights w_i in (5) satisfy $w_i = \lambda(1-\lambda)^{i-1}$, $i = 1, 2,\ldots$, we get the AE formula (4). If the weights are

$$w_1 = 1 + \beta,\; w_2 = -\beta,\; w_i = 0 \text{ for } i = 3, 4, \ldots,$$

we get

$$\pi^e_{t-1,t} = (1+\beta)\pi_{t-1} - \beta\pi_{t-2} = \pi_{t-1} + \beta(\pi_{t-1} - \pi_{t-2}). \tag{6}$$

This is called the hypothesis of *extrapolative expectations* and says:

if $\beta > 0$, then the recent direction of change in π is expected to continue;

if $\beta < 0$, then the recent direction of change in π is expected to be reversed;

if $\beta = 0$, then expectations are static as in (2).

As hinted, there *are* cases where for instance myopic expectations *are* "rational" (in a sense to be defined below). Exercise 1 provides an example. But in many cases purely backward-looking formulas are too rigid, too mechanistic. They will often lead to systematic expectation errors to one side or the other. It seems implausible that people should not then respond to their experience and revise their expectations formula. When expectations are about things that really matter for them, people are likely to listen to professional forecasters who build their forecasting on statistical or econometric *models*. Such models are based on a formal probabilistic framework, take the interaction between different variables into account, and incorporate new information about future possible events.

2 The rational expectations hypothesis

2.1 Preliminaries

We first recapitulate a few concepts from statistics. A sequence $\{X_t\}$ of random variables indexed by time is called a *stochastic process.* A stochastic process $\{X_t\}$ is called *white noise* if for all t, X_t has zero expected value, constant variance, and zero covariance across time.[1] A stochastic process $\{X_t\}$ is called a *first-order autoregressive process,* abbreviated AR(1), if $X_t = \beta_0 + \beta_1 X_{t-1} + \varepsilon_t$, where β_0 and β_1 are constants, and $\{\varepsilon_t\}$ is white noise.

[1] The expression white noise derives from electrotechnics. In electrotechnical systems signals will often be subject to noise. If this noise is arbitrary and has no dominating frequence, it looks like white light. The various colours correspond to a certain wave length, but white light is light which has all frequences (no dominating frequence).

If $|\beta_1| < 1$, then $\{X_t\}$ is called a *stationary* AR(1) process. A stochastic process $\{X_t\}$ is called a *random walk* if $X_t = X_{t-1} + \varepsilon_t$, where $\{\varepsilon_t\}$ is white noise.

Before defining the term rational expectation, it is useful to clarify a distinction between two ways in which expectations, whatever their nature, may enter a macroeconomic model.

2.1.1 Two model types

Type A: models with past expectations of current endogenous variables

Suppose a given macroeconomic model can be reduced to two equations, the first being

$$Y_t = a\, Y_{t-1,t}^e + c\, X_t, \qquad t = 0, 1, 2, ..., \tag{7}$$

where Y_t is some endogenous variable (not necessarily GDP), a and c are given constant coefficients, and X_t is an exogenous random variable which follows some specified stochastic process. In line with the notation from Section 1, $Y_{t-1,t}^e$ is the subjective expectation formed in period $t-1$, of the value of the variable Y in period t. The economic agents are in simple models assumed to have the same expectations. Or, at least there is a dominating expectation, $Y_{t-1,t}^e$, in the society. What the equation (7) claims is that the endogenous variable, Y_t, depends, in the specified linear way, on the "generally held" expectation of Y_t, formed in the previous period. It is natural to think of the outcome Y_t as being the aggregate result of agents' decisions and market mechanisms, the decisions being made at discrete points in time $\ldots, t-2, t-1, t, \ldots$, immediately after the uncertainty concerning the period in question is resolved.

The second equation specifies how the subjective expectation is formed. To fix ideas, let us assume myopic expectations,

$$Y_{t-1,t}^e = Y_{t-1}, \tag{8}$$

as in (2) above. A *solution* to the model is a stochastic process for Y_t such that (7) holds, given the expectation formation (8) and the stochastic process which X_t follows.

EXAMPLE 1 *(imported raw materials and the domestic price level)* Let the endogenous variable in (7) represent the domestic price level (the consumer price index) P_t, and let X_t be the price level of imported raw materials. Suppose the price level is determined through a markup on unit costs,

$$P_t = (\lambda W_t + \eta X_t)(1 + \mu), \qquad 0 < \lambda < \frac{1}{1+\mu}, \tag{*}$$

where W_t is the nominal wage level in period $t = 0, 1, 2, \ldots$, and λ and η are positive technical coefficients representing the assumed constant labor and raw materials requirements, respectively, per unit of output; μ is a constant markup. Assume further that workers in period $t-1$ negotiate next period's wage level, W_t, so as to achieve, in expected value, a certain target real wage which we normalize to 1, i.e.,

$$\frac{W_t}{P^e_{t-1,t}} = 1.$$

Inserting into (*), we have

$$P_t = a\ P^e_{t-1,t} + c\ X_t, \qquad 0 < \alpha = \lambda(1+\mu) < 1, 0 < c = \eta(1+\mu). \tag{9}$$

Suppose $X_t = \bar{x} + \varepsilon_t$, where $\bar{x}$ is a positive constant and $\{\varepsilon_t\}$ is white noise. Assuming myopic expectations,

$$P^e_{t-1,t} = P_{t-1}, \tag{10}$$

the solution for the evolution of the price level is

$$P_t = a\ P_{t-1} + c(\bar{x} + \varepsilon_t), \qquad t = 0, 1, 2, \ldots.$$

Without shocks, and starting from an arbitrary $P_{-1} > 0$, the time path of the price level would be $P_t = (P_{-1} - P^*)a^{t+1} + P^*$, where $P^* = c\bar{x}/(1-\alpha)$. Shocks to the price of imported raw materials result in transitory deviations from P^*. But as the shocks are only temporary and $|a| < 1$, the domestic price level gradually returns towards the constant level P^*. The intervening changes in wage demands in response to the changes in the price level changes prolong the time it takes to return to P^* in the absence of new shocks. □

Equation (7) can also be interpreted as a vector equation (such that Y_t and $Y^e_{t-1,t}$ are n-vectors, a is an $n \times n$ matrix, c an $n \times m$ matrix, and X an m-vector). The crucial feature is that the endogenous variables dated t *only* depend on previous expectations of date-t values of these variables and on the exogenous variables.

Models with past expectations of current endogenous variables will serve as our point of reference when introducing the concept of rational expectations below.

Type B: models with forward-looking expectations

Another way in which agents' expectations may enter is exemplified by

$$Y_t = a\ Y^e_{t,t+1} + c\ X_t, \qquad t = 0, 1, 2, ... \tag{11}$$

Here $Y^e_{t,t+1}$ is the subjective expectation, formed in period t, of the value of Y in period $t+1$. Example: the equity price today depends on what the equity price is expected to be tomorrow. Or more generally: the current expectation of a future value of an endogenous variable influences the current value of this variable. We name this the case of *forward-looking expectations*. (In "everyday language" also $Y^e_{t-1,t}$ in model type 1 can be said to be a forward-looking variable as seen from period $t-1$. But the dividing line between the two model types, (7) and (11), is whether *current* expectations of future values of the endogenous variables do or do not influence the current values of these.)

The complete model with forward-looking expectations will include an additional equation, specifying how the subjective expectation, $Y^e_{t,t+1}$, is formed. We might again impose myopic expectations, $Y^e_{t,t+1} = Y_t$. A *solution* to the model is a stochastic process for Y_t satisfying (11), given the stochastic process followed by X_t and given the specified expectation formation and perhaps some additional restrictions in the form of boundary conditions or similar. The case of forward-looking expectations is important in connection with many topics in macroeconomics, including the evolution of asset prices and issues of asset price bubbles. This case will be dealt with in sections 3 and 4 below.

In passing we note that in both model type 1 and model type 2, it is the mean (in the subjective probability distribution) of the random variable(s) that enters. This is typical of simple macroeconomic models which often ignore other measures such as the median, mode, or higher-order moments. The latter, say the variance of X_t, may be included in more advanced models where for instance behavior towards risk is important.

2.1.2 The concept of a model-consistent expectation

The concepts of a *rational expectation* and *model-consistent expectation* are closely related, but not the same. We start with the latter.

Let there be given a stochastic model represented by (7) combined with some given expectation formation (8), say. We put ourselves in the position of the investigator or model builder and ask what the *model-consistent expectation* of the endogenous variable Y_t is as seen from period $t-1$. It is the mathematical *conditional expectation* that can be calculated on the basis of the model and available relevant data revealed up to and including period $t-1$. Let us denote this expectation

$$E(Y_t|I_{t-1}), \tag{12}$$

where E is the expectation operator and I_{t-1} denotes the information available at time $t-1$. We think of period $t-1$ as the half-open time interval $[t-1, t)$ and imagine that the uncertainty concerning the exogenous random variable X_{t-1} is resolved at time $t-1$.

So I_{t-1} includes knowledge of X_{t-1} and thereby, via the model, also of Y_{t-1}.

The information I_{t-1} may comprise knowledge of the realized values of X and Y up until and including period $t-1$. Instead of (12) we could, for instance, write

$$E(Y_t|Y_{t-1}=y_{t-1},\ldots,Y_{t-n}=y_{t-n};X_{t-1}=x_{t-1},\ldots,X_{t-n}=x_{t-n}).$$

Here information (some of which may be redundant) goes back to a given initial period, say period 0, in which case n equals t. Alternatively, perhaps information goes back to "ancient times", possibly represented by $n=\infty$. Anyway, as time proceeds, in general more and more realizations of the exogenous and endogenous variables become known and in *this* sense the information I_{t-1} *expands* with rising t. The information I_{t-1} may also be interpreted as "partial lack of uncertainty", so that an "increasing amount of information" and "reduced uncertainty" are seen as two sides of the same thing. The "reduced uncertainty" lies in the fact that the space of *possible* time paths $\{(X_t,Y_t)\}_{t-n}^{t+T}$ as of time t *shrinks* as time proceeds (T denotes the time horizon as seen from time t).[2] Indeed, this space shrinks precisely because more and more realizations of the variables take place (more information appears) and thereby rule out an increasing subset of paths that were earlier possible.[3]

In Example 1, as long as the subjective expectation is the myopic expectation (10), the model-consistent expectation is

$$E(P_t|I_{t-1})=a\ P_{t-1}+c\bar{x}.$$

Inserting the investigator's estimated values of the coefficients a and c, the investigator's forecast of P_t is obtained.

2.2 The rational expectations hypothesis

Unsatisfied with mechanistic formulas like those of Section 1, the American economist John F. Muth (1961) introduced a radically different approach, the hypothesis of *rational expectations.* Muth stated the hypothesis the following way:

> I should like to suggest that expectations, since they are informed predictions of future events, are essentially the same as the predictions of the relevant

[2]By "possible" is meant "ex ante feasible according to a given model".

[3]We refer to I_{t-1} as the "available information" rather than the "information set" which is an alternative term used in the literature. The latter term is tricky because, as we have just exemplified, it is ambiguous what is meant by a "larger information set". Moreover, the term "information set" has different meanings in different branches of economics, hence we are hesitant to use it. More about subtleties relating to "information" in Appendix B, dealing with mathematical conditional expectations in general.

economic theory. At the risk of confusing this purely descriptive hypothesis with a pronouncement as to what firms ought to do, we call such expectations 'rational' (Muth 1961).

Muth applied this hypothesis to simple microeconomic problems. The hypothesis was subsequently extended and applied to general equilibrium theory and macroeconomics by what since the early 1970s became known as the New Classical Macroeconomics school. Nobel laureate Robert E. Lucas from the University of Chicago lead the way by a series of papers starting with Lucas (1972) and Lucas (1973). Assuming rational expectations in a model instead of, for instance, adaptive expectations may radically change the dynamics as well as the impact of economic policy.

2.2.1 The concept of rational expectations

Assuming the economic agents have *rational expectations* (RE) is to assume that their subjective expectation equals the model-consistent expectation, that is, the mathematical conditional expectation that can be calculated on the basis of the model and available relevant information about the exogenous stochastic variables. In connection with the model ingredient (7), assuming the agents have rational expectations thus means that

$$Y_{t-1,t}^{e} = E(Y_t | I_{t-1}), \tag{13}$$

i.e., agents' subjective conditional expectation coincides with the "objective" or "true" conditional expectation, given the model (7).

Together, the equations (7) and (13) constitute a simple *rational expectations model* (henceforth an RE model). We may write the model in compact form as

$$Y_t = aE(Y_t | I_{t-1}) + c\ X_t, \qquad t = 0, 1, 2, ... \tag{14}$$

The assumption of rational expectations thus relies on idealized conditions.

2.2.2 Solving a simple RE model

To solve the model means to find the stochastic process followed by Y_t, given the stochastic process followed by the exogenous variable X_t. For a linear RE model with past expectations of current endogenous variables, the solution procedure is the following.

1. By substitution, reduce the RE model (or the relevant part of the model) into a form like (14) expressing the endogenous variable in period t in terms of its past

expectation and the exogenous variable(s). (The case with multiple endogenous variables is treated similarly.)

2. Take the conditional expectation on both sides of the equation and solve for the conditional expectation of the endogenous variable.

3. Insert into the "reduced form" attained at 1.

In practice there is often a fourth step, namely to express *other* endogenous variables in the model in terms of those found in step 3. Let us see how the procedure works by way of the following example.

EXAMPLE 2 We modify Example 1 by replacing myopic expectations by rational expectations, i.e., (10) is replaced by $P^e_{t-1,t} = E(P_t|I_{t-1})$. Now "available information" includes that the subjective expectations are rational expectations. Step 1:

$$P_t = aE(P_t|I_{t-1}) + c\ X_t, \qquad 0 < \alpha < 1, c > 0. \tag{15}$$

Step 2: $E(P_t|I_{t-1}) = aE(P_t|I_{t-1}) + c\bar{x}$, implying

$$E(P_t\,|I_{t-1}) = c\frac{\bar{x}}{1-a}.$$

Step 3: Insert into (15) to get

$$P_t = c\frac{a\bar{x}}{1-a} + c(\bar{x} + \varepsilon_t).$$

This is the solution of the model in the sense of a specification of the stochastic process followed by P_t.

To compare with myopic expectations, suppose the event $\varepsilon_t \neq 0$ is relatively seldom and that at $t = 0,\ 1,\ ...,\ t_0 - 1$, it so happens that $\varepsilon_t = 0$, hence $P_t = c\bar{x}/(1-a) \equiv P^*$. Then, at $t = t_0$, $\varepsilon_{t_0} > 0$, so that $P_{t_0} = P^* + c\varepsilon_{t_0} > P^*$. But for $t = t_0 + 1,\ t_0 + 2,\ ...,\ t_0 + n$ there is again a sequence of periods with $\varepsilon_t = 0$. Then, under RE, domestic price level returns to P^* already in period $t_0 + 1$.

With myopic expectations, combined with $P_{-1} = P^*$, say, the positive shock to import prices at $t = t_0$ will imply $P_{t_0} = aP^* + c(\bar{x} + \varepsilon_{t_0}) = P^* + c\varepsilon_{t_0}$, $P_{t_0+1} = a(P^* + c\varepsilon_t) + c\bar{x}$ $= P^* + ac\varepsilon_t$, $P_{t_0+i} = P^* + a^i c\varepsilon_t$ for $i = 1, 2,\ ..., n$. After t_0 there is a systematic positive forecast error. This is because the mechanical expectation does not consider how the economy really functions. □

Returning to the general form (14), without specifying the process $\{X_t\}$, the second step gives

$$E(Y_t\,|I_{t-1}) = c\frac{E(X_t\,|I_{t-1})}{1-a}, \tag{16}$$

when $a \neq 1$.[4] Then, in the third step we get

$$Y_t = c\frac{aE(X_t\,|I_{t-1}) + (1-a)X_t}{1-a} = c\frac{X_t - a(X_t - E(X_t\,|I_{t-1}))}{1-a}. \tag{17}$$

For instance, let X_t follow the process $X_t = \bar{x} + \rho X_{t-1} + \varepsilon_t$, where $0 < \rho < 1$ and ε_t has zero expected value, given all observed past values of X and Y. Then (17) yields the solution

$$Y_t = c\frac{X_t - a\varepsilon_t}{1-a} = c\frac{\bar{x} + \rho X_{t-1} + (1-a)\varepsilon_t}{1-a}, \qquad t = 0, 1, 2, ...$$

In Exercise 2 you are asked to solve a simple Keynesian model of this form and compare the solution under rational expectations with the solution under static expectations.

Rational expectations should be viewed as a simplifying assumption that at best offers an approximation. *First*, the assumption entails essentially that the economic agents share one and the same understanding about how the economic system functions (and in this chapter they also share one and the same information, I_{t-1}). This is already a big mouthful. *Second*, this perception is assumed to *comply with the model* of the informed economic specialist. *Third*, this model is supposed to be the *true* model of the economic process, including the true parameter values as well as the true stochastic process which X_t follows. Indeed, by equalizing $Y^e_{t-1,t}$ with the true conditional expectation, $E(Y_t|I_{t-1})$, and not at most some econometric estimate of this, it is presumed that agents know the true values of the parameters a and c in the data-generating process which the model is supposed to mimic. In practice it is not possible to attain such precise knowledge, at least not unless the considered economic system has reached some kind of steady state and no structural changes occur (a condition which is hardly ever satisfied in macroeconomics).

Nevertheless, a model based on the rational expectations hypothesis can in many contexts be seen as a useful cultivation of a theoretical research question. The results that emerge cannot be due to *systematic* expectation errors from the economic agents' side. In this sense the assumption of rational expectations makes up a theoretically interesting *benchmark case*.

We shall stick to the term "rational expectation" because it is standard. The term can easily be misunderstood, however. Usually, in economists' terminology "rational"

[4] If $a = 1$, the model (14) is inconsistent unless $E(X_t\,|I_{t-1})) = 0$ in which case there are multiple solutions. Indeed, for any number $k \in (-\infty, +\infty)$, the process $Y_t = k + cX_t$ solves the model when $E(X_t\,|I_{t-1}) = 0$.

refers to behavior based on optimization subject to the constraints faced by the agent. So one might think that the RE hypothesis stipulates that economic agents try to get the most out of a situation with limited information, contemplating the benefits and costs of gathering more information and using adequate statistical estimation methods. But this is a misunderstanding. The RE hypothesis presumes that the true model is already known to the agents. The "rationality" refers to taking this assumed knowledge fully into account in the chosen actions.

2.2.3 The forecast error*

Let the forecast of some variable Y one period ahead be denoted $Y^e_{t-1,t}$. Suppose the forecast is determined by some given function, f , of realizations of Y and X up to and including period $t-1$, that is, $Y^e_{t-1,t} = f(y_{t-1}, y_{t-2}, ..., x_{t-1}, x_{t-2}, ...)$. Such a function is known as a *forecast function*. It might for instance be one of the mechanistic forecasting principles in Section 1. At the other extreme the forecast function might, at least theoretically, coincide with the a model-consistent conditional expectation. In the latter case it is a *model-consistent forecast function* and we can write

$$
\begin{aligned}
& f(y_{t-1}, y_{t-2}, ..., x_{t-1}, x_{t-2}, ...) = E(Y_t \,| I_{t-1}) \qquad (18) \\
= \;& E(Y_t \,| Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}, ..., x_{t-1} = x_{t-1}, x_{t-2} = x_{t-2}, ...) \,.
\end{aligned}
$$

The *forecast error* is the difference between the actually occurring future value, Y_t, of a variable and the forecasted value. So, for a given forecast, $Y^e_{t-1,t}$, the *forecast error* is $e_t \equiv Y_t - Y^e_{t-1,t}$ and is itself a stochastic variable.

If the forecast function in (18) complies with the true data-generating process (a big "if"), then the implied forecasts would have several ideal properties:

(a) the forecast error would have zero mean;

(b) the forecast error would be uncorrelated with any of the variables in the information I_{t-1} and therefore also with its own past values; and

(c) the expected squared forecast error would be minimized.

To see these properties, note that the model-consistent forecast error is $e_t = Y_t - E(Y_t \,| I_{t-1})$. From this follows that $E(e_t \,| I_{t-1}) = 0$, cf. (a). Also the unconditional expectation is nil, i.e., $E(e_t) = 0$. This is because $E(E(e_t \,| I_{t-1})) = E(0) = 0$ at the same time as

$E(E(e_t \,|I_{t-1})) = E(e_t)$, by the *law of iterated expectations* from statistics saying that the unconditional expectation of the conditional expectation of a stochastic variable Z is given by the unconditional expectation of Z, cf. Appendix B. Considering the specific model (7), the model-consistent-forecast error is $e_t = Y_t - E(Y_t \,|I_{t-1}) = c(X_t - E(X_t \,|I_{t-1}))$, by (16) and (17). An ex post error ($e_t \neq 0$) thus emerges if and only if the realization of the exogenous variable deviates from its conditional expectation as seen from the previous period.

As to property (b), for $i = 1, 2, ...$, let s_{t-i} be some variable value belonging to the information I_{t-i}. Then, property (b) is the claim that the (unconditional) covariance between e_t and s_{t-i} is zero, i.e., $\text{Cov}(e_t s_{t-i}) = 0$, for $i = 1, 2, ...$ This follows from the *orthogonality property* of model-consistent expectations (see Appendix C). In particular, with $s_{t-i} = e_{t-i}$, we get $\text{Cov}(e_t e_{t-i}) = 0$, i.e., the forecast errors exhibit *lack of serial correlation.* If the covariance were not zero, it would be possible to improve the forecast by incorporating the correlation into the forecast. In other words, under the assumption of rational expectations economic agents have no more to learn from past forecast errors. As remarked above, the RE hypothesis precisely refers to a fictional situation where learning has been completed and underlying mechanisms do not change.

Finally, a desirable property of a forecast function $f(\cdot)$ is that it maximizes "accuracy", i.e., minimizes an appropriate loss function. A popular loss function, L, in this context is the expected squared forecast error conditional on the information I_{t-1},

$$L = E((Y_t - f(y_{t-1}, y_{t-2}, ..., x_{t-1}, x_{t-2}, ...))^2 \,|I_{t-1}) \,.$$

Assuming $Y_t, Y_{t-1}, ..., X_{t-1}, X_{t-2}, ...$ are jointly normally distributed, then the solution to the problem of minimizing L is to set $f(\cdot)$ equal to the conditional expectation $E(Y_t \,|I_{t-1})$ based on the data-generating model as in (18).[5] This is what property (c) refers to.

EXAMPLE 3 Let $Y_t = aE(Y_t \,|I_{t-1}) + cX_t$, with $X_t = \bar{x} + \varepsilon_t$, where $\bar{x}$ is a constant and ε_t is white noise with variance σ^2. Then (17) applies, so that

$$Y_t = \frac{c\bar{x}}{1-a} + c\varepsilon_t, \qquad t = 0, 1, ...,$$

with variance $c^2\sigma^2$. The model-consistent forecast error is $e_t = Y_t - E(Y_t \,|I_{t-1}) = c\varepsilon_t$ with conditional expectation equal to $E(c\varepsilon_t \,|I_{t-1}) = 0$. This forecast error itself is white noise and is therefore uncorrelated with the information on which the forecast is based. □

[5] For proof, see Pesaran (1987). Under the restriction of only *linear* forecast functions, property (c) holds even without the joint normality assumption, see Sargent (1979).

It is worth emphasizing that the "true" conditional expectation usually can not be known − neither to the economic agents nor to the investigator. At best there can be a reasonable estimate, probably somewhat different across the agents because of differences in information and conceptions of how the economic system functions. A deeper model of expectations would give an account of the mechanisms through which agents *learn* about the economic environment. An important ingredient here would be how agents contemplate the costs and potential gains associated with further information search needed to reduce systematic expectation errors where possible. This contemplation is intricate because information search often means entering unknown territory. Moreover, for a significant subset of the agents the costs may be prohibitive. A further complicating factor involved in learning is that when the agents have obtained *some* knowledge about the statistical properties of the economic variables, the resulting behavior of the agents may *change* these statistical properties. The rational expectations hypothesis sets these problems aside. It is simply assumed that the structure of the economy remains unchanged and that the learning process has been completed.

2.2.4 Perfect foresight as a special case

The notion of *perfect foresight* corresponds to the limiting case where the variance of the exogenous variable(s) is zero so that with probability one, $X_t = E(X_t \,|I_{t-1})$ for all t. Then we have a non-stochastic model where rational expectations imply that agents' ex post forecast error with respect to Y_t is zero.[6] To put it differently: rational expectations in a non-stochastic model is equivalent to perfect foresight. Note, however, that perfect foresight necessitates the exogenous variable X_t to be known in advance. Real-world situations are usually not like that. If we want our model to take this into account, the model ought to be formulated in an explicit stochastic framework. And assumptions should be stated about how the economic agents respond to the uncertainty. The rational expectations assumption is one approach to the problem and has been much applied in macroeconomics in recent decades, perhaps due to lack of compelling tractable alternatives.

3 Models with rational forward-looking expectations

We here turn to models where current expectations of a future value of an endogenous variable have an influence on the current value of this variable, that is, the case exemplified

[6] Here we disregard zero probability events.

by equation (11). At the same time we introduce two simplifications in the notation. First, instead of using capital letters to denote the stochastic variables (as we did above and is common in mathematical statistics), we follow the tradition in macroeconomics to often use lower case letters. So a lower case letter may from now on represent a stochastic variable *or* a specific value of this variable, depending on the context.

An equation like (11) will now read $y_t = a\ y^e_{t,t+1} + c\ x_t$. Under rational expectations it takes the form $y_t = aE(y_{t+1}\,|I_t) + c\ x_t$, $t = 0, 1, 2, \ldots$. Second, from now on we write this equation as

$$y_t = aE_t y_{t+1} + c\ x_t, \ldots t = 0, 1, 2, \ldots,\ \ a \neq 0. \tag{19}$$

That is, the expected value of a stochastic variable, z_{t+i}, conditional on the information I_t, will be denoted $E_t z_{t+i}$.

A stochastic difference equation of the form (19) is called a linear *expectation difference equation of first order* with constant coefficient a.[7] A *solution* is a specified stochastic process $\{y_t\}$ which satisfies (19), given the stochastic process followed by x_t. In the economic applications usually no initial value, y_0, is given. On the contrary, the interpretation is that y_t depends, for all t, on expectations about the future.[8] So y_t is considered a *jump variable* that can immediately shift its value in response to the emergence of new information about the future x's. For example, a share price may immediately jump to a new value when the accounts of the firm become publicly known (often even before, due to sudden rumors).

Due to the lack of an initial condition for y_t, there can easily be infinitely many processes for y_t satisfying our expectation difference equation. We have an infinite forward-looking "regress", where a variable's value today depends on its expected value tomorrow, this value depending on the expected value the day after tomorrow and so on. Then usually there are infinitely many expected sequences which can be self-fulfilling in the sense that if only the agents expect a particular sequence, then the aggregate outcome of their behavior will be that the sequence is realized. It "bites its own tail" so to speak. Yet, when an equation like (19) is part of a larger model, there will often (but not always) be conditions that allow us to select *one* of the many solutions to (19) as the only *economically* relevant one. For example, an economy-wide transversality condition or another general

[7] To keep things simple, we let the coefficients a and c be constants, but a generalization to time-dependent coefficients is straightforward.

[8] The reason we say "depends on" is that it would be inaccurate to say that y_t is *determined* (in a one-way-sense) by expectations about the future. Rather there is *mutual dependence.* In view of y_t being an element in the information I_t, the expectation of y_{t+1} in (19) may depend on y_t just as much as y_t depends on the expectation of y_{t+1}.

equilibrium condition may rule out divergent solutions and leave a unique convergent solution as the final solution.

We assume $a \neq 0$, since otherwise (19) itself is already the unique solution. It turns out that the set of solutions to (19) takes a different form depending on whether $|a| < 1$ or $|a| > 1$:

The case $|a| < 1$. In general, there is a unique *fundamental solution* and infinitely many explosive solutions ("bubble solutions").

The case $|a| > 1$. In general, there is no fundamental solution but infinitely many non-explosive solutions. (The case $|a| = 1$ resembles this.)

In the case $|a| < 1$, the expected future has modest influence on the present. Here we will concentrate on this case, since it is the case most frequently appearing in macroeconomic models with rational expectations.

4 Solutions when $|a| < 1$

Various solution methods are available. *Repeated forward substitution* is the most easily understood method.

4.1 Repeated forward substitution

Repeated forward substitution consists of the following steps. We first shift (19) one period ahead:

$$y_{t+1} = a\ E_{t+1}y_{t+2} + c\ x_{t+1}.$$

Then we take the conditional expectation on both sides to get

$$E_t y_{t+1} = a\ E_t(E_{t+1}y_{t+2}) + c\ E_t x_{t+1} = a\ E_t y_{t+2} + c\ E_t x_{t+1}, \tag{20}$$

where the second equality sign is due to the *law of iterated expectations*, which says that

$$E_t(E_{t+1}y_{t+2}) = E_t y_{t+2}. \tag{21}$$

see Box 1. Inserting (20) into (19) then gives

$$y_t = a^2 E_t y_{t+2} + ac\ E_t x_{t+1} + c\ x_t. \tag{22}$$

The procedure is repeated by forwarding (19) two periods ahead; then taking the conditional expectation and inserting into (22), we get

$$y_t = a^3 E_t y_{t+3} + a^2 c\ E_t x_{t+2} + ac\ E_t x_{t+1} + c\ x_t.$$

We continue in this way and the general form (for $n = 0, 1, 2, ...$) becomes

$$y_{t+n} = a\ E_{t+n}(y_{t+n+1}) + c\ x_{t+n},$$

$$E_t y_{t+n} = a\ E_t y_{t+n+1} + c\ E_t x_{t+n},$$

$$y_t = a^{n+1} E_t y_{t+n+1} + c x_t + c \sum_{i=1}^{n} a^i E_t x_{t+i}. \tag{23}$$

Box 1. The law of iterated expectations

The method of repeated forward substitution is based on the law of iterated expectations which says that $E_t(E_{t+1} y_{t+2}) = E_t y_{t+2}$, as in (21). The logic is the following. Events in period $t+1$ are stochastic as seen from period t and so $E_{t+1} y_{t+2}$ (the expectation conditional on these events) is a stochastic variable. Then the law of iterated expectations says that the conditional expectation of this stochastic variable as seen from period t is the same as the conditional expectation of y_{t+2} itself as seen from period t. So, given that expectations are rational, then an earlier expectation of a later expectation of y is just the earlier expectation of y. Put differently: my best forecast today of how I am going to forecast tomorrow a share price the day after tomorrow, will be the same as my best forecast today of the share price the day after tomorrow. If beforehand we have good reasons to expect that we will revise our expectations upward, say, when next period's additional information arrives, the original expectation would be biased, hence not rational.[9]

4.2 The fundamental solution

PROPOSITION 1 Consider the expectation difference equation (19), where $a \neq 0$. If

$$\lim_{n \to \infty} \sum_{i=1}^{n} a^i E_t x_{t+i} \text{ exists,} \tag{24}$$

then

$$y_t = c \sum_{i=0}^{\infty} a^i E_t x_{t+i} = c x_t + c \sum_{i=1}^{\infty} a^i E_t x_{t+i} \equiv y_t^*, \qquad t = 0, 1, 2, ..., \tag{25}$$

is a solution to the equation.

[9] A formal account of conditional expectations and the law of iterated expectations is given in Appendix B.

Proof Assume (24). Then the formula (25) is meaningful. In view of (23), it satisfies (19) if and only if $\lim_{n\to\infty} a^{n+1} E_t y_{t+n+1} = 0$. Hence, it is enough to show that the process (25) satisfies this latter condition.

In (25), replace t by $t+n+1$ to get $y_{t+n+1} = c\sum_{i=0}^{\infty} a^i E_{t+n+1} x_{t+n+1+i}$. Using the law of iterated expectations, this yields

$$\begin{aligned} E_t y_{t+n+1} &= c\sum_{i=0}^{\infty} a^i E_t x_{t+n+1+i} \qquad \text{so that} \\ a^{n+1} E_t y_{t+n+1} &= c\, a^{n+1} \sum_{i=0}^{\infty} a^i E_t x_{t+n+1+i} = c \sum_{j=n+1}^{\infty} a^j E_t x_{t+j}. \end{aligned}$$

It remains to show that $\lim_{n\to\infty} \sum_{j=n+1}^{\infty} a^j E_t x_{t+j} = 0$. From the identity

$$\sum_{j=1}^{\infty} a^j E_t x_{t+j} = \sum_{j=1}^{n} a^j E_t x_{t+j} + \sum_{j=n+1}^{\infty} a^j E_t x_{t+j}$$

follows

$$\sum_{j=n+1}^{\infty} a^j E_t x_{t+j} = \sum_{j=1}^{\infty} a^j E_t x_{t+j} - \sum_{j=1}^{n} a^j E_t x_{t+j}.$$

Letting $n \to \infty$, this gives

$$\lim_{n\to\infty} \sum_{j=n+1}^{\infty} a^j E_t x_{t+j} = \sum_{j=1}^{\infty} a^j E_t x_{t+j} - \sum_{j=1}^{\infty} a^j E_t x_{t+j} = 0,$$

which was to be proved. □

The solution (25) is called the *fundamental solution* of (19), often marked by an asterisk *. The fundamental solution is (for $c \neq 0$) defined only when the condition (24) holds. In general this condition requires that $|a| < 1$. In addition, (24) requires that the absolute value of the expectation of the exogenous variable does not increase "too fast". More precisely, the requirement is that $|E_t x_{t+i}|$, when $i \to \infty$, has a growth factor less than $|a|^{-1}$. As an example, let $0 < a < 1$ and $g > 0$, and suppose that $E_t x_{t+i} > 0$ for $i = 0, 1, 2, ...$, and that $1 + g$ is an upper bound for the growth factor of $E_t x_{t+i}$. Then

$$E_t x_{t+i} \leq (1+g) E_t x_{t+i-1} \leq (1+g)^i E_t x_t = (1+g)^i x_t.$$

Multiplying by a^i, we get $a^i E_t x_{t+i} \leq a^i (1+g)^i x_t$. By summing from $i = 1$ to n,

$$\sum_{i=1}^{n} a^i E_t x_{t+i} \leq x_t \sum_{i=1}^{n} [a(1+g)]^i .$$

Letting $n \to \infty$, we get

$$\lim_{n\to\infty} \sum_{i=1}^{n} a^i E_t x_{t+i} \leq x_t \lim_{n\to\infty} \sum_{i=1}^{n} [a(1+g)]^i = x_t \frac{a(1+g)}{1-a(1+g)} < \infty,$$

if $1 + g < a^{-1}$, using the sum rule for an infinite geometric series.

As noted in the proof of Proposition 1, the fundamental solution, (25), has the property that

$$\lim_{n\to\infty} a^n E_t y_{t+n} = 0. \tag{26}$$

That is, the expected value of y is not "explosive": its absolute value has a growth factor less than $|a|^{-1}$. Given $|a| < 1$, the fundamental solution is the only solution of (19) with this property. Indeed, it is seen from (23) that whenever (26) holds, (25) must also hold. In Example 1 below, y_t is interpreted as the market price of a share and x_t as dividends. Then the fundamental solution gives the share price as the present value of the expected future flow of dividends.

EXAMPLE 1 *(the fundamental value of an equity share)* Consider arbitrage between shares of stock and a riskless asset paying the constant rate of return $r > 0$. Let period t be the current period. Let p_{t+i} be the market price (in real terms, say) of the share at the beginning of period $t + i$ and d_{t+i} the dividend paid out at the end of that period, $t + i$, $i = 0, 1, 2, ...$ As seen from period t there is uncertainty about p_{t+i} and d_{t+i} for $i = 1, 2, ...$ An investor who buys n_t shares at time t (the beginning of period t) thus invests $V_t \equiv p_t n_t$ units of account at time t. At the end of the period the gross return comes out as the known dividend $d_t n_t$ and the potential sales value of the shares at the beginning of next period. This is unlike standard *accounting* and *finance* notation in discrete time, where V_t would be the end-of-period-t market value of the stock of shares that begins to yield dividends in period $t + 1$.[10]

Suppose investors have rational expectations and care only about expected return.

[10] Our use of p_t for the (real) price of a share bought at the beginning of period t is not inconsistent with our use, in earlier chapters, of P_t to denote the nominal price per unit of consumption in period t, but paid for at the *end* of the period. At the beginning of period t, after the uncertainty pertaining to period t has been resolved and available information thereby been updated, a consumer-investor will decide both the investment and the consumption flow for the period. But only the investment expence, p_t, is disbursed immediately.

It is convenient to think of the course of actions such that receipt of the previous period's dividend, d_{t-1}, and payment for that period's consumption, at the price P_{t-1}, occur right before period t begins and the new information arrives. Indeed, the resolution of uncertainty at discrete points in time motivates a *distinction* between "end of" period $t - 1$ and "beginning of" period t, where the new information has just arrived.

Then the no-arbitrage condition reads

$$\frac{d_t + E_t p_{t+1} - p_t}{p_t} = r > 0. \tag{27}$$

This can be written

$$p_t = \frac{1}{1+r} E_t p_{t+1} + \frac{1}{1+r} d_t, \tag{28}$$

which is of the same form as (19) with $a = c = 1/(1+r) \in (0, 1)$. Assuming dividends do not grow "too fast", we find the fundamental solution, denoted p_t^*, as

$$p_t^* = \frac{1}{1+r} d_t + \frac{1}{1+r} \sum_{i=1}^{\infty} \frac{1}{(1+r)^i} E_t d_{t+i} = \sum_{i=0}^{\infty} \frac{1}{(1+r)^{i+1}} E_t d_{t+i}. \tag{29}$$

The fundamental solution is simply the present value of expected future dividends.

If the dividend process is $d_{t+1} = d_t + \varepsilon_{t+1}$, where ε_{t+1} is white noise, then the dividend process is known as a *random walk* and $E_t d_{t+i} = d_t$ for $i = 1, 2, \ldots$. Thus $p_t^* = d_t/r$, by the sum rule for an infinite geometric series. In this case the fundamental value is thus itself a random walk. More generally, the dividend process could be a *martingale,* that is, a sequence of stochastic variables with the property that the expected value next period exists and equals the current actual value, i.e., $E_t d_{t+1} = d_t$; but in a martingale, $\varepsilon_{t+1} \equiv d_{t+1} - d_t$ need not be white noise; it is enough that $E_t \varepsilon_{t+1} = 0$.[11] Given the constant required return r, we still have $p_t^* = d_t/r$. So the fundamental value itself is in this case a martingale. □

In finance theory the present value of the expected future flow of dividends on an equity share is referred to as the *fundamental value* of the share. It is by analogy with this that the general designation *fundamental solution* has been introduced for solutions of form (25). We could also think of p_t as the market price of a house rented out and d_t as the rent. Or p_t could be the market price of an oil well and d_t the revenue (net of extraction costs) from the extracted oil in period t.

4.3 Bubble solutions

Other than the fundamental solution, the expectation difference equation (19) has infinitely many *bubble solutions.* In view of $|a| < 1$, these are characterized by violating the condition (26). That is, they are solutions whose expected value explodes over time.

[11] A random walk is thus a special case of a martingale.

It is convenient to first consider the *homogenous* expectation equation associated with (19). This is defined as the equation emerging when setting $c = 0$ in (19):

$$y_t = aE_t y_{t+1}. \tag{30}$$

Every stochastic process $\{b_t\}$ of the form

$$b_{t+1} = a^{-1}b_t + u_{t+1}, \quad \text{where } E_t u_{t+1} = 0, \tag{31}$$

has the property that

$$b_t = aE_t b_{t+1}, \tag{32}$$

and is thus a solution to (30). The "disturbance" u_{t+1} represents "new information" which may be related to movements in "fundamentals", x_{t+1}. But it does not have to. In fact, u_{t+1} may be related to conditions that *per se* have no economic relevance whatsoever.

For ease of notation, from now on we just write b_t even if we think of the whole process $\{b_t\}$ rather than the value taken by b in the specific period t. The meaning should be clear from the context. A solution to (30) is referred to as a *homogenous solution* associated with (19). Let b_t be a given homogenous solution and let K be an arbitrary constant. Then $B_t = Kb_t$ is also a homogenous solution (try it out for yourself). Conversely, any homogenous solution b_t associated with (19) can be written in the form (31). To see this, let b_t be a given homogenous solution, that is, $b_t = aE_t b_{t+1}$. Let $u_{t+1} = b_{t+1} - E_t b_{t+1}$. Then

$$b_{t+1} = E_t b_{t+1} + u_{t+1} = a^{-1}b_t + u_{t+1},$$

where $E_t u_{t+1} = E_t b_{t+1} - E_t b_{t+1} = 0$. Thus, b_t is of the form (31).

For convenience we here repeat our original expectation difference equation (19) and name it (*):

$$y_t = aE_t y_{t+1} + c\, x_t, \ldots t = 0, 1, 2, \ldots, \quad a \neq 0. \tag{*}$$

PROPOSITION 2 Consider the expectation difference equation (*), where $a \neq 0$. Let $\tilde{y}_t$ be a particular solution to the equation. Then:

(i) every stochastic process of the form

$$y_t = \tilde{y}_t + b_t, \tag{33}$$

where b_t satisfies (31), is a solution to (*);

(ii) every solution to (*) can be written in the form (33) with b_t being an appropriately chosen homogenous solution associated with (*).

Proof. Let some particular solution $\tilde{y}_t$ be given. (i) Consider $y_t = \tilde{y}_t + b_t$, where b_t satisfies (31). Since $\tilde{y}_t$ satisfies (*), we have $y_t = a\ E_t\tilde{y}_{t+1} + c\ x_t + b_t$. Consequently, by (30),

$$y_t = a\ E_t\tilde{y}_{t+1} + c\ x_t + a\ E_t b_{t+1} = a\ E_t(\tilde{y}_{t+1} + b_{t+1}) + c\ x_t = a\ E_t y_{t+1} + c\ x_t,$$

saying that (33) satisfies (*). (ii) Let Y_t be an arbitrary solution to (*). Define $b_t = Y_t - \tilde{y}_t$. Then we have

$$\begin{aligned} b_t &= Y_t - \tilde{y}_t = aE_tY_{t+1} + cx_t - (aE_t\tilde{y}_{t+1} + cx_t) \\ &= aE_t(Y_{t+1} - \tilde{y}_{t+1}) = aE_t b_{t+1}, \end{aligned}$$

where the second equality follows from the fact that both Y_t and $\tilde{y}_t$ are solutions to (*). This shows that b_t is a solution to the homogenous equation (30) associated with (*). Since $Y_t = \tilde{y}_t + b_t$, the proposition is hereby proved. □

Proposition 2 holds for any $a \neq 0$. In case the fundamental solution (25) exists and $|a| < 1$, it is convenient to choose this solution as the particular solution in (33). Thus, referring to the right-hand side of (25) as y_t^*, we can use the particular form,

$$y_t = y_t^* + b_t. \tag{34}$$

When the component b_t is different from zero, the solution (34) is called a *bubble solution* and b_t is called the *bubble component*. In the typical economic interpretation the bubble component shows up only because it is expected to show up next period, cf. (32). The name bubble springs from the fact that the expected value of b_t, conditional on the information available in period t, explodes over time when $|a| < 1$. To see this, as an example, let $0 < a < 1$. Then, from (30), by repeated forward substitution we get

$$b_t = a\ E_t(aE_{t+1}b_{t+2}) = a^2E_tb_{t+2} = ... = a^iE_tb_{t+i}, \quad i = 1, 2, ...$$

It follows that $E_tb_{t+i} = a^{-i}b_t$, and from this follows that the bubble, for t going to infinity, is unbounded in expected value:

$$\lim_{i\to\infty} E_tb_{t+i} = \begin{cases} \infty, \text{ if } b_t > 0 \\ -\infty, \text{ if } b_t < 0 \end{cases}. \tag{35}$$

Indeed, the absolute value of E_tb_{t+i} will for rising i grow *geometrically* towards infinity with a growth factor equal to $1/a > 1$.

Let us consider a special case of (*) that allows a simple graphical illustration of both the fundamental solution and some bubble solutions.

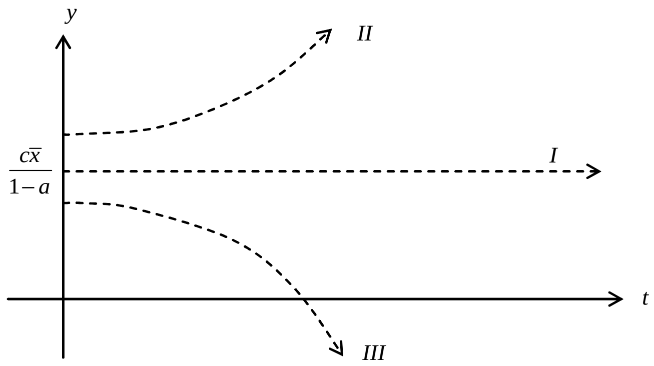

Figure 1: Deterministic bubbles (the case $0 < a < 1$, $c > 0$, and $x_t = \bar{x}$).

4.3.1 When x_t has constant mean

Suppose the stochastic process x_t (the "fundamentals") takes the form $x_t = \bar{x} + \varepsilon_t$, where $\bar{x}$ is a constant and ε_t is white noise. Then

$$y_t = a\ E_t y_{t+1} + c(\bar{x} + \varepsilon_t), \qquad 0 < |a| < 1. \tag{36}$$

The fundamental solution is

$$y_t^* = c\ x_t + c\sum_{i=1}^{\infty} a^i \bar{x} = c\bar{x} + c\varepsilon_t + c\frac{a\bar{x}}{1-a} = \frac{c\bar{x}}{1-a} + c\varepsilon_t.$$

Referring to (i) of Proposition 2,

$$y_t = \frac{c\bar{x}}{1-a} + c\varepsilon_t + b_t \tag{37}$$

is thus also a solution of (36) if b_t is of the form (31).

It may be instructive to consider the case where all stochastic features are eliminated. So we assume $u_t \equiv \varepsilon_t \equiv 0$. Then we have a model with perfect foresight; the solution (37) simplifies to

$$y_t = \frac{c\bar{x}}{1-a} + b_0 a^{-t}, \tag{38}$$

where we have used repeated *backward* substitution in (31). By setting $t = 0$ we see that $y_0 - \frac{c\bar{x}}{1-a} = b_0$. Inserting this into (38) gives

$$y_t = \frac{c\bar{x}}{1-a} + (y_0 - \frac{c\bar{x}}{1-a})a^{-t}. \tag{39}$$

In Fig. 1 we have drawn three trajectories for the case $0 < a < 1$, $c > 0$. Trajectory I has $y_0 = c\bar{x}/(1-a)$ and represents the fundamental solution. Trajectory II, with $y_0 > c\bar{x}/(1-a)$, and trajectory III, with $y_0 < c\bar{x}/(1-a)$, are bubble solutions. Since we have imposed no boundary condition apriori, one y_0 is as good as any other. The interpretation is that there are infinitely many trajectories with the property that if only the economic agents expect the economy will follow that particular trajectory, the aggregate outcome of their behavior will be that this trajectory is realized. This is the potential indeterminacy arising when y_t is not a predetermined variable. However, as alluded to above, in a complete economic model there will often be restrictions on the endogenous variable(s) not visible in the basic expectation difference equation(s), here (36). It may be that the economic meaning of y_t precludes negative values (a share certificate would be an example). In that case no-one can rationally expect a path such as III in Fig. 1. Or perhaps, for some reason, there is an upper bound on y_t (think of the full-employment ceiling for output in a situation where the "natural" growth factor for output is smaller than a^{-1}). Then no one can rationally expect a trajectory like II in the figure.

To sum up: in order for a solution of a first-order linear expectation difference equation with constant coefficient a, where $|a| < 1$, to differ from the fundamental solution, the solution must have the form (34) where b_t has the form described in (31). This provides a clue as to what asset price bubbles might look like.

4.3.2 Asset price bubbles

A stylized fact of stock markets is that stock price indices are quite volatile on a month-to-month, year-to-year, and especially decade-to-decade scale, cf. Fig. 2. There are different views about how these swings should be understood. According to the *Efficient Market Hypothesis* the swings just reflect unpredictable changes in the "fundamentals", that is, changes in the present value of rationally expected future dividends. This is for instance the view of Nobel laureate Eugene Fama (1970, 2003) from University of Chicago.

In contrast, Nobel laureate Robert Shiller (1981, 2003, 2005) from Yale University, and others, have pointed to the phenomenon of "excess volatility". The view is that asset prices tend to fluctuate more than can be rationalized by shifts in information about fundamentals (present values of dividends). Although in no way a verification, graphs like those in Fig. 2 and Fig. 3 are suggestive. Fig. 2 shows the monthly real Standard and Poors (S&P) composite stock prices and real S&P composite earnings for the period 1871-2008. The unusually large increase in real stock prices since the mid-90's, which ended with the collapse in 2000, is known as the "dot-com bubble". Fig. 3 shows, on a monthly basis, the ratio of real S&P stock prices to an average of the previous ten years' real S&P earnings along with the long-term real interest rate. It is seen that this ratio

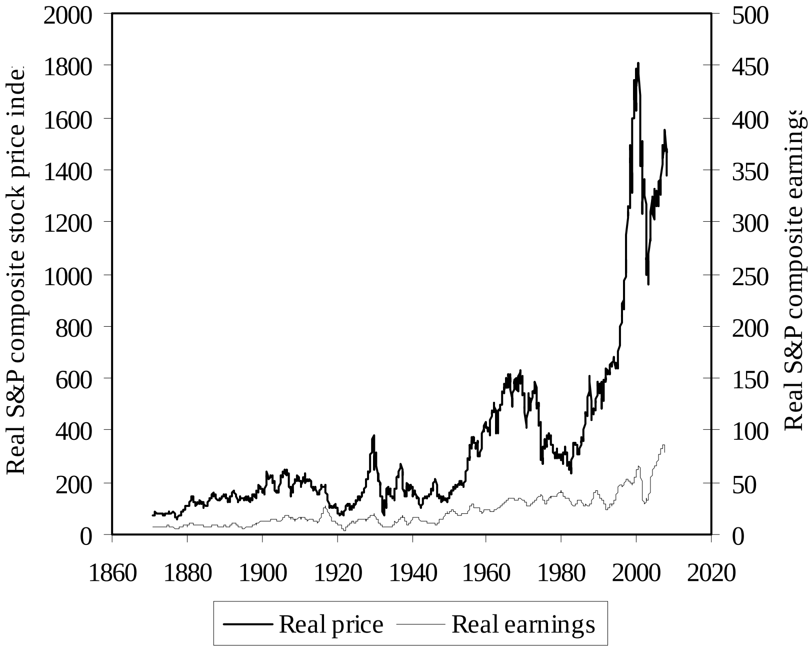

Figure 2: Monthly real S&P composite stock prices from January 1871 to January 2008 (left) and monthly real S&P composite earnings from January 1871 to September 2007 (right). Source: http://www.econ.yale.edu/~shiller/data.htm.

reached an all-time high in 2000, by many observers considered as "the year the dot-com bubble burst".

Shiller's interpretation of the large stock market swings is that they are due to fads, herding, and shifts in fashions and "animal spirits" (the latter being a notion from Keynes).

A third possible source of large stock market swings was pointed out by Blanchard (1979) and Blanchard and Watson (1982). They argued that bubble phenomena need not be due to irrational behavior and non-rational expectations. This lead to the theory of *rational bubbles* – the idea that excess volatility can be explained as speculative bubbles arising from self-fulfilling *rational* expectations.

Consider an asset which yields either dividends or services in production or consumption in every period in the future. The fundamental value of the asset is, at the theoretical level, defined as the present value of the expected future flow of dividends or services.[12] An *asset price bubble* is then defined as a systematic positive deviation of the market

[12] In practice there are many ambiguities involved in this definition of the fundamental value because it relates to a future which is often essentially unknown.

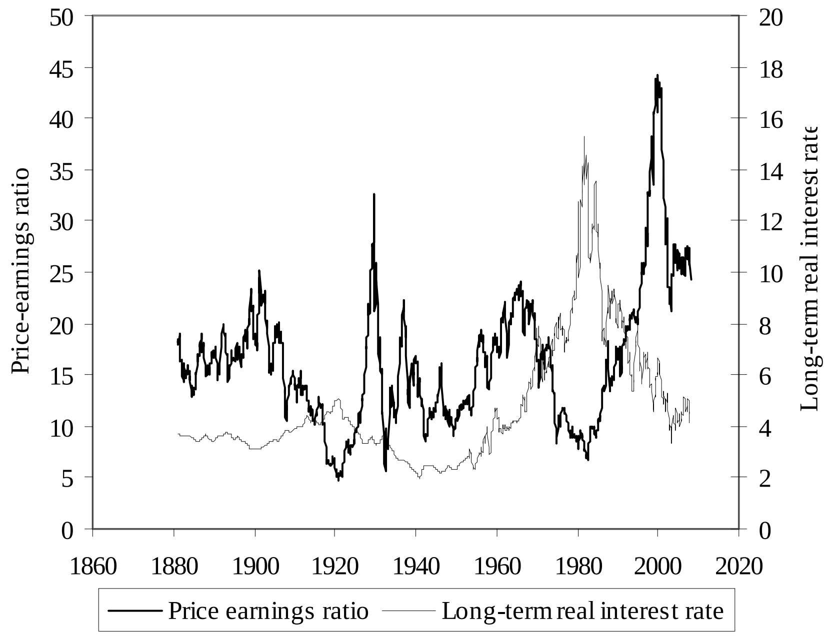

Figure 3: S&P price-earnings ratio and long-term real interest rates from January 1881 to January 2008. The earnings are calculated as a moving average over the preceding ten years. The long-term real interest rate is the 10-year Treasury rate from 1953 and government bond yields from Sidney Homer, "A History of Interest Rates" from before 1953. Source: http://www.econ.yale.edu/~shiller/data.htm.

price, p_t, of the asset from its fundamental value, p_t^*:

$$p_t = p_t^* + b_t. \tag{40}$$

An asset price bubble, $p_t - p_t^*$, that emerges in a setting where the no-arbitrage condition (27) holds under rational expectations, is called a *rational bubble*. It emerges only because there is in the market a self-fulfilling belief that it will appreciate at a rate high enough to warrant the overcharge involved.

EXAMPLE 2 *(an ever-expanding rational bubble)* Consider again an equity share for which the no-arbitrage condition is

$$\frac{d_t + E_t p_{t+1} - p_t}{p_t} = r > 0. \tag{41}$$

As in Example 1, the implied expectation difference equation is $p_t = aE_t p_{t+1} + cd_t$, with $a = c = 1/(1+r) \in (0,1)$. Let the price of the share at time t be $p_t = p_t^* + b_t$, where p_t^* is the fundamental value and $b_t > 0$ a bubble component following the deterministic process, $b_{t+1} = (1+r)b_t$, $b_0 > 0$, so that $b_t = b_0(1+r)^t$. This is called a *deterministic rational bubble*. The sum $p_t^* + b_t$ will satisfy the no-arbitrage condition (41) just as much as p_t^* itself, because we just add something which equals the discounted value of itself one period later.

Agents may be ready to pay a price over and above the fundamental value (whether or not they know the "true" fundamental value) if they expect they can sell at a sufficiently higher price later; trading with such motivation is called *speculative behavior*. If generally held and lasting for some time, this expectation may be self-fulfilling. Note that (41) implies that the asset price ultimately grows at the rate r. Indeed, let $d_t = d_0(1+\gamma)^t$, $\gamma < r$ (if $r \leq \gamma$, the asset price would be infinite). By the rule of the sum of an infinite geometric series, we then have $p_t^* = d_t/(r-\gamma)$, showing that the fundamental value grows at the rate γ. Consequently, $p_t/b_t = (p_t^* + b_t)/b_t = p_t^*/b_t + 1 \to 1$, as $\gamma < r$. It follows that the asset price in the long run grows at the same rate as the bubble, the rate r.

We are not acquainted with *ever*-expanding incidents of that caliber in real world situations, however. A deterministic rational bubble thus appears implausible. □

In some contexts it may not matter whether or not we think of the "rational" market participants as actually knowing the probability distribution of the "fundamentals", hence knowing p_t^* (by "fundamentals" is meant any information relating to the future dividend or service capacity of an asset: a firm's technology, resources, market conditions etc.). All the same, it seems common to imply such a high level of information in the term "rational bubbles". Unless otherwise indicated, we shall let this implication be understood.

While a deterministic rational bubble was found implausible, let us now consider an example of a *stochastic* rational bubble which sooner or later *bursts.*

EXAMPLE 3 *(a bursting bubble)* Once again we consider the no-arbitrage condition is (41) where for simplicity we still assume the required rate of return is constant, though possibly including a risk premium. Following Blanchard (1979), we assume that the market price, p_t, of the share contains a stochastic bubble of the following form:

$$b_{t+1} = \begin{cases} \frac{1+r}{q_t} b_t & \text{with probability } q_t, \\ 0 & \text{with probability } 1 - q_t, \end{cases} \tag{42}$$

where $t = 0, 1, 2, ...$ and $b_0 > 0$. In addition we may assume that $q_t = f(p_t^*, b_t)$, $f_{p^*} \geq 0$, $f_b \leq 0$. If $f_{p^*} > 0$, the probability that the bubble persists at least one period ahead is higher the greater the fundamental value has become. If $f_b < 0$, the probability that the bubble persists at least one period ahead is less, the greater the bubble has already become. In this way the probability of a crash becomes greater and greater as the share price comes further and further away from fundamentals. As a compensation, the longer time the bubble has lasted, the higher is the expected growth rate of the bubble in the absence of a collapse.

This bubble satisfies the criterion for a rational bubble. Indeed, (42) implies

$$E_t\ b_{t+1} = (\frac{1+r}{q_{t+1}}b_t)q_{t+1} + 0 \cdot (1-q_{t+1}) = (1+r)b_t.$$

This is of the form (31) with $a^{-1} = 1+r$, and the bubble is therefore a stochastic rational bubble. The stochastic component is $u_{t+1} = b_{t+1} - E_t b_{t+1} = b_{t+1} - (1+r)b_t$ and has conditional expectation equal to zero. Although u_{t+1} must have zero conditional expectation, it need not be white noise (it can for instance have varying variance). □

As this example illustrates, a stochastic rational bubble does not have the implausible ever-expanding form of a deterministic rational bubble. Yet, under certain conditions even stochastic rational bubbles can be ruled out or at least be judged implausible. The next section reviews some cases.

4.4 When rational bubbles in asset prices can or can not be ruled out

We concentrate on assets whose services are valued independently of the price.[13] Let p_t be the market price and p_t^* the fundamental value of the asset as of time t. Even if the asset yields services rather than dividends, we think of p_t^* as in principle the same for all agents. This is because a user who, in a given period, values the service flow of the asset relatively low can hire it out to the one who values it highest (the one with the highest willingness to pay). Until further notice we assume p_t^* known to the market participants.

4.4.1 Partial equilibrium arguments

The principle of reasoning to be used is called *backward induction:* If we know something about an asset price in the future, we can conclude something about the asset price today.

(a) Assets which can be freely disposed of ("free disposal") Can a rational asset price bubble be *negative*? The answer is no. The logic can be illustrated on the basis of Example 2 above. For simplicity, let the dividend be the same constant $d > 0$ for all $t = 0, 1, 2, ...$ Then, from the formula (39) we have

$$p_t - p^* = (p_0 - p^*)(1+r)^t,$$

[13] This is in contrast to assets that serve as means of payment.

where $r > 0$ and $p^* = d/r$. Suppose there is a negative bubble in period 0, i.e., $p_0 - p^* < 0$. In period 1, since $1 + r > 1$, the bubble is greater in absolute value. The downward movement of p_t continues and sooner or later p_t is negative. The intuition is that the low p_0 in period 0 implies a high dividend-price ratio. Hence a negative capital gain ($p_{t+1} - p_t < 0$) is needed for the no-arbitrage condition (41) to hold. Thereby $p_1 < p_0$, and so on.

But in a market with self-interested rational agents, an object which can be freely disposed of can never have a negative price. A negative price means that the "seller" has to *pay* to dispose of the object. Nobody will do that if the object can just be thrown away. An asset which can be freely disposed of (share certificates for instance) can therefore never have a negative price. We conclude that a *negative* rational bubble can not be consistent with rational expectations. Similarly, with a stochastic dividend, a negative rational bubble would imply that in expected value the share price becomes negative at some point in time, cf. (35). Again, rational expectations rule this out.

Hence, if we imagine that for a short moment $p_t < p_t^*$, then everyone will want to *buy* the asset and hold it forever, which by own use or by hiring out will imply a discounted value equal to p_t^*. There is thus excess demand until p_t has risen to p_t^*.

When a negative rational bubble can be ruled out, then, if at the first date of trading of the asset there were no positive bubble, neither can a positive bubble arise later. Let us make this precise:

PROPOSITION 3 Assume free disposal of a given asset. Then, if a rational bubble in the asset price is present today, it must be positive and must have been present also yesterday and so on back to the first date of trading the asset. And if a rational bubble bursts, it will not restart later.

Proof As argued above, in view of free disposal, a negative rational bubble in the asset price can be ruled out. It follows that $b_t = p_t - p_t^* \geq 0$ for $t = 0, 1, 2, ...$, where $t = 0$ is the first date of trading the asset. That is, any rational bubble in the asset price must be a positive bubble. We now show by contradiction that if, for an arbitrary $t = 1, 2, ...$, it holds that $b_t > 0$, then $b_{t-1} > 0$. Let $b_t > 0$. Then, if $b_{t-1} = 0$, we have $E_{t-1}b_t = E_{t-1}u_t$ $= 0$ (from (31) with t replaced by $t-1$), implying, since $b_t < 0$ is not possible, that $b_t = 0$ with probability *one* as seen from period $t-1$. Ignoring zero probability events, this rules out $b_t > 0$ and we have arrived at a contradiction. Thus $b_{t-1} > 0$. Replacing t by $t-1$ and so on backward in time, we end up with $b_0 > 0$. This reasoning also implies that if a bubble bursts in period t, it can not restart in period $t+1$, nor, by extension, in any subsequent period. □

This proposition (due to Diba and Grossman, 1988) claims that a rational bubble in an asset price must have been there since trading of the asset began. Yet such a conclusion is not without ambiguities. If new information about radically new technology comes up at some point in time, is a share in the firm then the same asset as before? In a legal sense the firm is the same, but is the asset also the same? Even if an earlier bubble has crashed, cannot a new rational bubble arise later in case of an utterly new situation?

These ambiguities reflect the difficulty involved in the concepts of rational expectations and rational bubbles when we are dealing with uncertainties about future developments of the economy. The market's evaluation of many assets of macroeconomic importance, not the least shares in firms, depends on vague beliefs about future preferences, technologies, and societal circumstances. The fundamental value can not be determined in any objective way. There is no well-defined probability distribution over the potential future outcomes. *Fundamental uncertainty*, also called *Knightian uncertainty*,[14] is present.

(b) Bonds with finite maturity The finite maturity ensures that the value of the bond is given at some finite future date. Therefore, if there were a positive bubble in the market price of the bond, no rational investor would buy just before that date. Anticipating this, no one would buy the date before, and so on. Consequently, nobody will buy in the first place. By this backward-induction argument follows that a positive bubble cannot get started. And since there also is "free disposal", *all* rational bubbles can be precluded.

From now on we take as given that negative rational bubbles are ruled out. So, the discussion is about whether *positive* rational asset price bubbles may exist or not.

(c) Assets whose supply is elastic Real capital goods (including buildings) can be reproduced and have clearly defined costs of reproduction. This precludes rational bubbles on this kind of assets, since a potential buyer can avoid the overcharge by producing instead. Notice, however, that building sites with a specific amenity value and apartments in attractive quarters of a city are not easily reproducible. Therefore, rational bubbles on such assets are more difficult to rule out.

[14] After the Chicago of University economist Frank Knight who in his book, *Risk, Uncertainty, and Profit* (1921), coined the important distinction between *measurable risk* and *unmeasurable uncertainty.*

Here are a few intuitive remarks about bubbles on shares of stock in an established firm. An argument against a rational bubble might be that if there were a bubble, the firm would tend to exploit it by issuing more shares. But thereby market participants mistrust is raised and may pull market evaluation back to the fundamental value. On the other hand, the firm might anticipate this adverse response from the market. So the firm chooses instead to "fool" the market by steady financing behavior, calmly enjoying its solid equity and continuing as if no bubble were present. It is therefore not obvious that this kind of argument can rule out rational bubbles on shares of stock.

(d) Assets for which there exists a "backstop-technology" For some articles of trade there exists substitutes in elastic supply which will be demanded if the price of the article becomes sufficiently high. Such a substitute is called a "backstop-technology". For example oil and other fossil fuels will, when their prices become sufficiently high, be subject to intense competition from substitutes (renewable energy sources). This precludes an unbounded bubble process in the price of oil.

On account of the arguments (c) and (d), it seems more difficult to rule out rational bubbles when it comes to assets which are not reproducible or substitutable, let alone assets whose fundamentals are difficult to ascertain. For some assets the fundamentals are not easily ascertained. Examples are paintings of past great artists, rare stamps, diamonds, gold etc. Also new firms that introduce completely novel products and technologies are potential candidates. Think of the proliferation of radio broadcasting in the 1920s before the wall Street crash in 1929 and the internet in the 1990s before the dotcom bubble burst in 2000.

What these situations allow for may not be termed rational bubbles, if by definition this concept requires a well-defined fundamental. Then we may think of a broader class of real-world bubbly phenomena driven by self-reinforcing expectations.

4.4.2 Adding general equilibrium arguments

The above considerations are of a partial equilibrium nature. On top of this, *general equilibrium* arguments can be put forward to limit the possibility of rational bubbles. We may briefly give a flavour of two such general equilibrium arguments. We still consider assets whose services are valued independently of the price and which, as in (a) above, can be freely disposed of. A house, a machine, or a share in a firm yields a service in consumption or production or in the form of a dividend stream. Since such an asset has an intrinsic value, p_t^*, equal to the present value of the flow of services, one might believe that positive rational bubbles on such assets can be ruled out in general equilibrium. As we shall see, this is indeed true for an economy with a finite number of "neoclassical" households (to be defined below), but not necessarily in an overlapping generations model. Yet even there, rational bubbles can under certain conditions be ruled out.

(e) An economy with a finite number of infinitely-lived households Assume that the economy consists of a finite number of infinitely-lived agents − here called households − indexed $i = 1, 2, ..., N$. The households are "neoclassical" in the sense that they save only with a view to future consumption.

Under free disposal in point (a) we saw that $p_t < p_t^*$ can not be an equilibrium. We now consider the case of a positive bubble, i.e., $p_t > p_t^*$. All owners of the bubble asset who are users will in this case prefer to *sell* and then *rent*; this would imply excess supply and could thus not be an equilibrium. Hence, we turn to households that are not users, but speculators. Assuming "short selling" is legal, speculators may pursue "short selling", that is, they first rent the asset (for a contracted interval of time) and immediately sell it at p_t. This results in excess supply and so the asset price falls towards p_t^*. Within the contracted interval of time the speculators buy the asset back and return it to the original owners in accordance with the loan accord. So $p_t > p_t^*$ can not be an equilibrium.

Even ruling out "short selling" (which *is* sometimes outright forbidden), we can exclude positive bubbles in the present setup with a finite number of households. To assume that owners who are not users would want to hold the bubble asset forever as a permanent investment will contradict that these owners are "neoclassical". Indeed, their transversality condition would be violated because the value of their wealth would grow at a rate asymptotically equal to the rate of interest. This would allow them to increase their consumption now without decreasing it later and without violating their No-Ponzi-Game condition.

We have to instead imagine that the "neoclassical" households who own the bubble asset, hold it against future sale. This could on the face of it seem rational enough if there were some probability that not only would the bubble continue to exist, but it would also grow so that the return would be at least as high as that yielded on an alternative investment. Owners holding the asset in the expectation of a capital gain, will thus plan to sell at some later point in time. Let t_i be the point in time where household

i wishes to sell and let

$$T = \max[t_1, t_2, ..., t_N].$$

Then nobody will plan to hold the asset after T. The household speculator, i, having $t_i = T$ will thus not have anyone to sell to (other than people who will only pay p_T^*). Anticipating this, no-one would buy or hold the asset the period before, and so on. So no-one will want to buy or hold the asset in the first place.

The conclusion is that $p_t > p_t^*$ cannot be a rational expectations equilibrium in a setup with a finite number of "neoclassical" households.

The same line of reasoning does not, however, go through in an overlapping generations model where *new* households – that is, new traders – enter the economy every period.

(f) An economy with interest rate above the output growth rate In an overlapping generations (OLG) model with an infinite sequence of new decision makers, rational bubbles are under certain conditions theoretically possible. The argument is that with $N \to \infty$, T as defined above is not bounded. Although this unboundedness is a necessary condition for rational bubbles, it is not sufficient, however.

To see why, let us return to the arbitrage examples 1, 2, and 3 where we have $a^{-1} = 1 + r$ so that a hypothetical rational bubble has the form $b_{t+1} = (1 + r)b_t + u_{t+1}$, where $E_t u_{t+1} = 0$. So in expected value the hypothetical bubble is growing at a rate equal to the interest rate, r. If at the same time r is higher than the long-run output growth rate, the value of the expanding bubble asset would sooner or later be larger than GDP and aggregate saving would not suffice to back its continued growth. Agents with rational expectations anticipate this and so the bubble never gets started.

This point is valid when the interest rate in the OLG economy is higher than the growth rate of the economy – which is normally considered the realistic case. Yet, the opposite case *is* possible and in that situation it is less easy to rule out rational asset price bubbles. This is also the case in situations with imperfect credit markets. It turns out that the presence of segmented financial markets or externalities that create a wedge between private and social returns on productive investment may increase the scope for rational bubbles (Blanchard, 2008).

4.5 Conclusion

The empirical evidence concerning asset price bubbles in general and rational asset price bubbles in particular seems inconclusive. It is very difficult to statistically distinguish between bubbles and mis-specified fundamentals. Rational bubbles can also have more complicated forms than the bursting bubble in Example 3 above. For example Evans (1991) and Hall et al. (1999) study "regime-switching" rational bubbles.

Whatever the possible limits to the plausibility of rational bubbles in asset prices, it is useful to be aware of their logical structure and the variety of forms they can take as logical possibilities. Rational bubbles may serve as a benchmark for a variety of "behavioral asset price bubbles", i.e., bubbles arising through particular psychological mechanisms. This would take us to *behavioral finance* theory. The reader is referred to, e.g., Shiller (2003).

For surveys on the theory of rational bubbles and econometric bubble tests, see Salge (1997) and Gürkaynak (2008). For discussions of famous historical bubble episodes, see the symposium in *Journal of Economic Perspectives* 4, No. 2, 1990, and Shiller (2005).

5 Appendix

A. The log-linear specification

In many macroeconomic models with rational expectations the equations are specified as log-linear, that is, as being linear in the logarithms of the variables. If Y, X, and Z are the original positive stochastic variables, defining $y = \ln Y$, $x = \ln X$, and $z = \ln Z$, a log-linear relationship between Y, X, and Z is a relation of the form

$$y = \alpha + \beta x + \gamma z, \tag{43}$$

where α, β, and γ are constants. The motivation for assuming log-linearity can be:

(a) Linearity is convenient because of the simple rule for the expected value of a sum: $E(\alpha+\beta x+\gamma z) = \alpha+\beta E(x)+\gamma E(z)$, where E is the expectation operator. Indeed, for a non-linear function, $f(x, z)$, we generally have $E(f(x, z)) \neq f(E(x), E(z))$.

(b) Linearity in logs may often seem a more realistic assumption than linearity in anything else.

(c) In time series models a logarithmic transformation of the variables followed by formation of first differences can be the road to eliminating a trend in the mean and variance.

As to point (b) we state the following:

CLAIM To assume linearity in logs is equivalent to assuming constant elasticities.

Proof Let the positive variables Y, X and Z be related by $Y = F(X,\ Z)$, where F is a continuous function with continuous partial derivatives. Taking the differential on both sides of $\ln\ Y = \ln F(X,Z)$, we get

$$\begin{aligned} d\ln Y &= \frac{1}{F(X,Z)}\frac{\partial F}{\partial X}dX + \frac{1}{F(X,Z)}\frac{\partial F}{\partial Z}dZ \\ &= \frac{X}{Y}\frac{\partial Y}{\partial X}\frac{dX}{X} + \frac{Z}{Y}\frac{\partial Y}{\partial Z}\frac{dZ}{Z} = \eta_{YX}\frac{dX}{X} + \eta_{YZ}\frac{dZ}{Z} = \eta_{YX}d\ln X + \eta_{YZ}d\ln Z, \end{aligned} \tag{44}$$

where η_{YX} and η_{YZ} are the partial elasticities of Y w.r.t. X and Z, respectively. Thus, defining $y = \ln Y$, $x = \ln X$, and $z = \ln Z$, gives

$$dy = \eta_{YX}dx + \eta_{YZ}dz. \tag{45}$$

Assuming constant elasticities amounts to putting $\eta_{YX} = \beta$ and $\eta_{YZ} = \gamma$, where β and γ are constants. Then we can write (45) as $dy = \beta dx + \gamma dz$. By integration, we get (43) where α is now an arbitrary integration constant. Hereby we have shown that constant elasticities imply a log-linear relationship between the variables.

Now, let us instead start by assuming the log-linear relationship (43). Then,

$$\frac{\partial y}{\partial x} = \beta, \frac{\partial y}{\partial z} = \gamma. \tag{46}$$

But (43), together with the definitions of y, x and z, implies that

$$Y = e^{\alpha+\beta x+\gamma z} = e^{\alpha+\beta\ln X+\gamma\ln Z},$$

from which follows that

$$\frac{\partial Y}{\partial X} = Y\beta\frac{1}{X} \text{ so that } \eta_{YX} \equiv \frac{X}{Y}\frac{\partial Y}{\partial X} = \beta,$$

and

$$\frac{\partial Y}{\partial Z} = Y\gamma\frac{1}{Z} \text{ so that } \eta_{YZ} \equiv \frac{Z}{Y}\frac{\partial Y}{\partial Z} = \gamma.$$

That is, the partial elasticities are constant. □

So, when the variables are in logs, then the coefficients in the linear expressions are the elasticities. Note, however, that the interest rate is normally an exception. It is often regarded as more realistic to let the interest rate itself and not its logarithm enter linearly. Then the associated coefficient indicates the *semi-elasticity* with respect to the interest rate.

B. Conditional expectations and the law of iterated expectations

The mathematical conditional expectation is a weighted sum of the possible values of the stochastic variable with weights equal to the corresponding conditional probabilities.

Let Y and X be two *discrete* stochastic variables with joint probability function $j(y,x)$ and marginal probability functions $f(y)$ and $g(x)$, respectively. If the conditional probability function for Y given $X = x_0$ is denoted $h(y\,|x_0)$, we have $h(y\,|x_0) = j(y,x_0)/g(x_0)$, assuming $g(x_0) > 0$. The conditional expectation of Y given $X = x_0$, denoted $E(Y|X = x_0)$, is then

$$E(Y|X = x_0) = \sum_y y\frac{j(y,x_0)}{g(x_0)}, \tag{47}$$

where the summation is over all the possible values of y.

This conditional expectation is a function of x_0. Since x_0 is just one possible value of the stochastic variable X, we interpret the conditional expectation itself as a stochastic variable and write it as $E(Y|X)$. Generally, for a function of the discrete stochastic variable X, say $k(X)$, the expected value is

$$E(k(X)) = \sum_x k(x)g(x).$$

When we here let the conditional expectation $E(Y|X)$ play the role of $k(X)$ and sum over all x for which $g(x) > 0$, we get

$$\begin{aligned} E(E(Y|X)) &= \sum_x E(Y|x)g(x) = \sum_x \left(\sum_y y\frac{j(y,x)}{g(x)}\right) g(x) \qquad \text{(by (47))} \\ &= \sum_y y\left(\sum_x j(y,x)\right) = \sum_y yf(y) = E(Y). \end{aligned}$$

This result is a manifestation of the *law of iterated expectations:* the unconditional expectation of the conditional expectation of Y is given by the unconditional expectation of Y.

Now consider the case where Y and X are *continuous* stochastic variables with joint probability *density* function $j(y,x)$ and marginal density functions $f(y)$ and $g(x)$, respectively. If the conditional density function for Y given $X = x_0$ is denoted $h(y\,|x_0)$, we have

$h(y\,|x_0) = j(y,x_0)/g(x_0)$, assuming $g(x_0) > 0$. The conditional expectation of Y given $X = x_0$ is

$$E(Y|X = x_0) = \int_{-\infty}^{\infty} y \frac{j(y,x_0)}{g(x_0)} dy, \tag{48}$$

where we have assumed that the range of Y is $(-\infty, \infty)$. Again, we may view the conditional expectation itself as a stochastic variable and write it as $E(Y|X)$. Generally, for a function of the continuous stochastic variable X, say $k(X)$, the expected value is

$$E(k(X)) = \int_R k(x)g(x)dx,$$

where R stands for the range of X. When we let the conditional expectation $E(Y|X)$ play the role of $k(X)$, we get

$$\begin{aligned} E(E(Y|X)) &= \int_R E(Y|x)g(x)dx = \int_R \left(\int_{-\infty}^{\infty} y \frac{j(y,x)}{g(x)} dy \right) g(x)dx \text{ (by (48))} \\ &= \int_{-\infty}^{\infty} y \left(\int_R j(y,x)dx \right) dy = \int_{-\infty}^{\infty} yf(y)dy = E(Y). \end{aligned} \tag{49}$$

This shows us the *law of iterated expectations* in action for continuous stochastic variables: the unconditional expectation of the conditional expectation of Y is given by the unconditional expectation of Y.

EXAMPLE Let the two stochastic variables, X and Y, follow a two-dimensional normal distribution. Then, from mathematical statistics we know that the conditional expectation of Y given X satisfies

$$E(Y|X) = E(Y) + \frac{\text{Cov}(Y,X)}{\text{Var}(X)}(X - E(X)).$$

Taking expectations on both sides gives

$$E(E(Y|X)) = E(Y) + \frac{\text{Cov}(Y,X)}{\text{Var}(X)}(E(X) - E(X)) = E(Y). \qquad \square$$

We may also express the law of iterated expectations in terms of subsets of the original outcome space for a stochastic variable. Let the event $\mathcal{A}$ be a subset of the outcome space for Y and let $\mathcal{B}$ be a subset of $\mathcal{A}$. Then the law of iterated expectations takes the form

$$E(E(Y|\mathcal{B})|\mathcal{A}) = E(Y|\mathcal{A}). \tag{50}$$

That is, when $\mathcal{B} \subseteq \mathcal{A}$, the expectation, conditional on $\mathcal{A}$, of the expectation of Y, conditional on $\mathcal{B}$, is the same as the expectation, conditional on $\mathcal{A}$, of Y.

Often we consider a dynamic context where expectations are conditional on dated information I_{t-i} $(i = 1, 2, ...)$. By a, so far, "informal analogy" with (49) we then write the law of iterated expectations this way:

$$E(E(Y_t|I_{t-i})) = E(Y_t), \qquad \text{for } i = 1, 2, ... \tag{51}$$

In words: the unconditional expectation of the conditional expectation of Y_t, given the information up to time $t - i$ equals the unconditional expectation of Y_t. Similarly, by a, so far, "informal analogy" with (50) we may write

$$E(E(Y_{t+2}|I_{t+1})|I_t) = E(Y_{t+2}|I_t). \tag{52}$$

That is, the expectation today of the expectation tomorrow, when more may be known, of a variable the day after tomorrow is the same as the expectation today of the variable the day after tomorrow. Intuitively: you ask a stockbroker in which direction she expects to revise her expectations upon the arrival of more information. If the broker answers "upward", say, then another broker is recommended.

The notation used in the transition from (50) to (52) might seem problematic, though. That is why we talk of "informal analogy". The sets $\mathcal{A}$ and $\mathcal{B}$ are subsets of the outcome space and $\mathcal{B} \subseteq \mathcal{A}$. In contrast, the "information" or "information content" represented by our symbol I_t will, for the uninitiated, inevitably be understood in a meaning *not* fitting the inclusion $I_{t+1} \subseteq I_t$. Intuitively "information" dictates the opposite inclusion, namely as a set which *expands* over time – more and more "information" (like "knowledge" or "available data") is revealed as time proceeds.

It is possible, however, to interpret the information I_t from another angle so as to make the notation in (52) fully comply with that in (50). Let the outcome space Ω denote the set of ex ante possible[15] sequences $\{(Y_t, X_t)\}_{t=t_0}^{T}$, where Y_t and X_t are vectors of date-t endogenous and exogenous stochastic variables, respectively, and where T is the time horizon, possibly $T = \infty$. For $t \in \{t_0, t_0 + 1, \ldots, T\}$, let the subset $\Omega_t \subseteq \Omega$ be defined as the of time t still possible sequences $\{(Y_s, X_s)\}_{s=t}^{T}$. Now, as time proceeds, more and more realizations occur, that is, more and more of the ex ante random states, (Y_t, X_t), become historical data, (y_t, x_t). Hence, as time proceeds, the subset Ω_t *shrinks* in the sense that $\Omega_{t+1} \subseteq \Omega_t$. The increasing amount of information and the "reduced uncertainty" can thus be seen as two sides of the same thing. Interpreting I_t this way, i.e., as "partial lack of uncertainty", the expression (52) means the same thing as

$$E(E(Y_{t+2}|\Omega_{t+1})|\Omega_t) = E(Y_{t+2}|\Omega_t).$$

[15] By "possible" is meant "feasible according to a given model".

This is in complete harmony with (50).

C. Properties of the model-consistent forecast

As in the text of Section 24.2.2, let e_t denote the model-consistent forecast error $Y_t - E(Y_t|I_{t-1})$. Then, if S_{t-1} represents information contained in I_{t-1},

$$\begin{aligned} E(e_t\,|S_{t-1}) &= E(Y_t - E(Y_t\,|I_{t-1})\,|S_{t-1}) = E(Y_t\,|S_{t-1}) - E(E(Y_t\,|I_{t-1})\,|S_{t-1}) \\ &= E(Y_t\,|S_{t-1}) - E(Y_t\,|S_{t-1}) = 0, \end{aligned} \tag{53}$$

where we have used that $E(E(Y_t\,|I_{t-1})\,|S_{t-1}) = E(Y_t\,|S_{t-1})$, by the law of iterated expectations. With $S_{t-1} = I_{t-1}$ we have, as a special case,

$$\begin{aligned} E(e_t\,|I_{t-1}) &= 0, \quad \text{as well as} \\ E(e_t) &= E(Y_t - E(Y_t\,|I_{t-1})) = E(Y_t) - E(E(Y_t\,|I_{t-1})) = 0, \end{aligned} \tag{54}$$

in view of (51) with $i = 1$. This proves property (a) in Section 24.2.3.

As to property (b) in Section 24.2.2, for $i = 1, 2, ...$, let s_{t-i} be an arbitrary variable value belonging to the information I_{t-i}. Then, $E(e_t s_{t-i}\,|I_{t-i}) = s_{t-i}E(e_t\,|I_{t-i}) = 0$, by (53) with $S_{t-1} = I_{t-i}$ (since I_{t-i} is contained in I_{t-1}). Thus, by the principle (51),

$$E(e_t s_{t-i}) = E\left(E(e_t s_{t-i}\,|I_{t-i})\right) = E(0) = 0 \quad \text{for } i = 1, 2, ... \tag{55}$$

This result is known as the *orthogonality property* of model-consistent expectations (two stochastic variables Z and V are said to be *orthogonal* if $E(ZV) = 0$). From the general formula for the (unconditional) covariance follows

$$\text{Cov}(e_t s_{t-i}) = E(e_t s_{t-i}) - E(e_t)E(s_{t-i}) = 0 - 0 = 0, \quad \text{for } i = 1, 2, ...,$$

by (54) and (55). In particular, with $s_{t-i} = e_{t-i}$, we get $\text{Cov}(e_t e_{t-i}) = 0$. This proves that model-consistent forecast errors exhibit *lack of serial correlation.*

6 Exercises

1. Let $\{X_t\}$ be a stochastic process in discrete time. Suppose $Y_t = X_t + e_t$ and $X_t = X_{t-1} + \varepsilon_t$, where e_t and ε_t are white noise.

a) Is $\{X_t\}$ a random walk? Why or why not?

b) Is $\{Y_t\}$ a random walk? Why or why not?

c) Calculate the rational expectation of X_t conditional on all relevant information up to and including period $t-1$.

d) What is the rational expectation of Y_t conditional on all relevant information up to and including period $t-1$?

e) Compare with the subjective expectation of Y_t based on the adaptive expectations formula with adjustment speed equal to one.

2. Consider a simple Keynesian model of a closed economy with constant wages and prices (behind the scene), abundant capacity, and output determined by demand:

$$Y_t = D_t = C_t + \bar{I} + G_t, \tag{1}$$

$$C_t = \alpha + \beta Y^e_{t-1,t}, \qquad \alpha > 0,\ 0 < \beta < 1, \tag{2}$$

$$G_t = (1-\rho)\bar{G} + \rho G_{t-1} + \varepsilon_t, \qquad \bar{G} > 0,\ 0 < \rho < 1, \tag{3}$$

where the endogenous variables are Y_t = output (= income), D_t = aggregate demand, C_t = consumption, and $Y^e_{t-1,t}$ = expected output (income) in period t as seen from period $t-1$, while G_t, which stands for government spending on goods and services, is considered exogenous as is ε_t, which is white noise. Finally, investment, $\bar{I}$, and the parameters α, β, ρ, and $\bar{G}$ are given positive constants.

Suppose expectations are "static" in the sense that expected income in period t equals actual income in the previous period.

a) Solve for Y_t.

b) Find the income multiplier (partial derivative of Y_t) with respect to a change in G_{t-1} and ε_t, respectively.

Suppose instead that expectations are rational.

c) Explain what this means.

d) Solve for Y_t.

e) Find the income multiplier with respect to a change in G_{t-1} and ε_t, respectively.

f) Compare the result under e) with that under b). Comment.

3. Consider arbitrage between equity shares and a riskless asset paying the constant rate of return $r > 0$. Let p_t denote the price at the beginning of period t of a share that at the end of period t yields the dividend d_t. As seen from period t there is uncertainty about p_{t+i} and d_{t+i} for $i = 1, 2, \ldots$. Suppose agents have rational expectations and care only about expected return (risk neutrality).

a) Write down the no-arbitrage condition.

Suppose dividends follow the process $d_t = \bar{d} + \varepsilon_t$, where $\bar{d}$ is a positive constant and ε_t is white noise, observable in period t, but not known in advance.

b) Find the fundamental solution for p_t and let it be denoted p_t^*. *Hint:* given $y_t = aE_t y_{t+1} + c\, x_t$, the fundamental solution is $y_t = cx_t + c\sum_{i=1}^{\infty} a^i E_t x_{t+i}$.

Suppose someone claims that the share price follows the process

$$p_t = p_t^* + b_t,$$

with a given $b_0 > 0$ and, for $t = 0, 1, 2, \ldots$,

$$b_{t+1} = \begin{cases} \frac{1+r}{q_t} b_t & \text{with probability } q_t, \\ 0 & \text{with probability } 1 - q_t, \end{cases}$$

where $q_t = f(b_t)$, $f' < 0$.

c) What is an asset price bubble and what is a rational asset price bubble?

d) Can the described b_t process be a rational asset price bubble? *Hint:* a bubble component associated with the inhomogenous equation $y_t = aE_t y_{t+1} + c\, x_t$ is a solution, different from zero, to the homogeneous equation, $y_t = aE_t y_{t+1}$.

—