14

Testing the Rational Expectations Hypothesis

In the remainder of this book, we consider the ways in which the RE hypothesis has been tested empirically. We begin in this chapter with the efficient market hypothesis, which is a joint hypothesis made up of rational expectations and a hypothesis about how expected returns are determined. In chapter 15, we discuss the evidence assembled by Robert Barro and others to test whether output is determined by the New Classical model or not. In chapter 16, we review direct tests of rational expectations using survey data on actual expectations; we then turn to the estimation of complete rational expectations models. Some such models are 'deep structure' (where the parameters are those of tastes and technology, presumably invariant to changes in the processes driving the exogenous variables including policy) but most are 'shallow structure', the parameters being those of aggregate supply and demand curves.

The purpose of this chapter is to consider the relationship between the concepts of financial market efficiency and rational expectations. A particular feature of financial markets is that trading can occur, in principle, almost continuously, and the market price is free to move to eradicate any imbalance between demand and supply. Furthermore since the assets traded (stocks, bonds, commodities) can be resold or traded in future periods it follows that financial markets are, more obviously than others, speculative in the technical sense that expectations of future asset prices affect current asset prices.

We begin this chapter by defining the concept of an efficient capital market and setting out models for the determination of normal or equilibrium asset returns and also the implications of rational expectations for the determination of asset prices. A brief review of the empirical evidence for efficiency in asset markets together with the problems faced in interpretation of the empirical evidence is also presented.

EFFICIENCY, PERFECTION AND ASSET PRICING IN CAPITAL MARKETS.

A capital or asset market is defined to be efficient when prices (e.g. stock prices, bond prices or exchange rates) fully and instantaneously reflect all available relevant information. Fama (1970) has defined three types of market efficiency, according to the extent of the information reflected in the market:

1. Weak-form efficient: A market is weak-form efficient if it is not possible for a trader to make abnormal returns by developing a trading rule based on the past history of prices or returns.

2. Semi-strong-form efficient: A market where a trader cannot make abnormal returns using a trading rule based on publicly available information. Examples of publicly available information are past money supply data, company financial accounts, or tipsters in periodicals.

3. Strong-form efficient: Where a trader cannot make abnormal returns using a trading rule based on any information source, whether public or private.

These three forms of efficiency represent a crude partitioning of all possible information systems into three broad categories, the precise boundaries of which are not easily defined. However, they are useful, as we shall see, for classifying empirical research on market efficiency. As their names suggest, strong-form efficiency implies semi-strong efficiency which in turn implies weak-form efficiency, while of course the reverse implications do not hold.

It is useful to distinguish between the concept of an efficient capital market and that of a perfect capital market. A perfect capital market could be defined as one in which the following conditions hold (see Copeland and Weston, 1988):

1. Markets are informationally efficient, that is information is costless and it is received simultaneously by individuals.

2. Markets are frictionless, that is there are no transactions costs or taxes, assets are perfectly divisible and marketable and there are no constraining regulations.

3. There is perfect competition in product and securities markets, that is agents are price-takers.

4. All individuals are rational expected-utility maximizers.

If conditions 1 to 4 were met (and assuming no significant distortions elsewhere in the economy), the capital market would be allocationally efficient, in that prices would be set to equate the marginal rates of return for all producers and savers, and of course consequently savings are optimally allocated. The notion of capital market efficiency is therefore much less restrictive. An element of imperfect competition in product markets would imply capital market imperfection; nevertheless, the stock market could determine a security price which fully reflected the present value of the stream of expected future monopoly profits. Consequently the stock market could still be efficient in the presence of imperfection.

Asset prices, in order to give the correct signals to traders, must fully and instantaneously reflect all available information. However, as pointed out by Grossman and Stiglitz (1976, 1980), it cannot be the case that market prices do fully and instantaneously reflect all available information. If this were so, agents would have no incentive for collecting and processing information, since it would already be reflected in the price, which each individual is assumed to be able to observe costlessly. It is the possibility of obtaining abnormal profits in the course of arbitraging which provides the incentive to collect and process new information. In the Grossman and Stiglitz model, individuals choose to become informed or remain uninformed, and in equilibrium each individual is indifferent between remaining uninformed on the one hand, and collecting information (or buying the expertise of brokers), so becoming informed, on the other. This is because after deducting information costs each action offers the same expected utility.[1]

---

[1] Hellwig (1982) challenges the Grossman and Stiglitz proposition that the informativeness of market prices in equilibrium is bounded away from full informational efficiency. Hellwig points out that this proposition rests on the assumption that agents learn from current prices at which transactions have actually been completed. This is a model in which investors learn from past equilibrium prices but not from the auctioneer's current price offer. Hellwig is able to show that if the time span between successive market transactions is short, the market can approximate full informational efficiency arbitrarily closely and yet the return to being informed remains bounded away from zero. This results from the fact that informed agents can utilize this information before uninformed agents have an opportunity to infer it from current market prices.

Hellwig also pursues the implications that arise if one relaxes the assumption that agents cannot assure themselves of being informed in a given period, but rather agents choose the frequency on average at which they obtain information. It appears that

Nevertheless, a reasonable interpretation of empirical tests of the efficient model hypothesis is, given that the data are collected at discrete intervals, that the process of arbitrage has occurred within the period. Consequently the implications of different available information can then be analysed without modelling the process of arbitrage itself (Begg, 1982) and this is the usual assumption in empirical work.

RATIONAL EXPECTATIONS AND MARKET EFFICIENCY

The semi-strong-form efficient markets model, i.e. that based on publicly available information, is an application of the concept of rational expectations, although this was not stressed in the early literature on efficiency, which goes back much further than the rational expectations literature. If expectations are non-rational, then publicly available information will not be reflected in asset prices and systematic abnormal profit opportunities will be available. This can be seen simply enough by noting that market agents have to know the model governing prices (or act as if they know it) in order to eliminate abnormal expected returns; if the model governing expected prices is different from that governing actual prices, there will be systematic abnormal returns available in the market.

Strong-form efficiency also implies that agents have rational expectations, since they must know how to use all sorts of private information as well as public; where strong-form efficiency differs from semi-strong is about the effects of private information on the market (fully discounted in strong-form, not at all in semi-strong).

Not quite the same is true of weak-form efficiency. In this case, since they make efficient use only of the past history of prices, they must know the time series model governing prices; strictly speaking this does not imply knowledge of the underlying structural model, since there will be generally insufficient identifying restrictions. Nevertheless, in practice, with limited samples and structural change, recovery of the time-series parameters by market agents from the data can effectively be ruled out. It is therefore natural, if not necessary, to assume in this case too that agents have rational expectations and so know the underlying model, from which they are then able to derive the time-series parameters.

While, therefore, market efficiency can be regarded as implying ratio-

---

relaxing such assumptions leaves his main result above unimpaired — see also the survey by Jordan and Radner (1982).

nal expectations, rational expectations does not imply market efficiency. Market efficiency is a joint hypothesis about expectations and market behaviour (specifically, the model of equilibrium expected returns). The main hypothesis about behaviour is the capital asset pricing model (see below); and in empirical tests more detailed assumptions must also be made about how equilibrium returns will move. Further hypotheses concern the behaviour of the agents with access to different sets of 'available' information. Under weak-form efficiency, active market participants are assumed to make effective use only of the past history of prices in their market; one theoretical basis for this has been in the costs of obtaining and processing wider information (Feige and Pierce, 1976). Under semi-strong-form efficiency, the assumption made is the usual one in rational expectations macro models that active agents use all publicly available information, presumed to be useable at zero or low cost. Finally, in the strong-form case it is assumed that those agents with access to private information deploy or indirectly influence funds to eliminate expected returns from this source of information; there are, however, problems with this, since private information cannot be sold at a fair price (once divulged it is valueless, but before divulgence it is impossible for the buyer to assess) and those with access have, by definition, in general only limited funds.

Because all definitions of market efficiency invoke the concept of abnormal returns, we are required for empirical work to have a theory of the equilibrium expected rate of return for assets. Tests of market efficiency are conducted after allowance for the equilibrium rate of return. If the riskiness of an asset does not change over time (or conversely if its risk changes randomly over time) then, for example, weak-form efficiency implies that there should not be an extrapolative pattern in the time series of returns. If there were a recurring pattern of any type, traders who recognized the pattern would use it to make abnormal profits. The very effort to use such patterns would, under the efficiency hypothesis, lead to their elimination.

THE CAPM MODEL OF EQUILIBRIUM EXPECTED RETURNS

The equilibrium expected return on assets is a central topic in modern portfolio theory and the interested reader is directed to, for example, Copeland and Weston (1988) for a full discussion (Allen, 1985, provides a helpful introduction). The following brief account of one such and widely used theory, the capital asset pricing model (CAPM), must suffice here.

Only if traders are risk-neutral[2] will they be indifferent to the variability of the returns (i.e. the risk) on their portfolio. Risk-averse individuals will be concerned about aggregate portfolio risk and will require a risk premium on each asset (or class of assets). By combining assets in the portfolio it is possible to diversify away some of the risk (the 'unsystematic' risk) associated with an asset. However, to the extent that the returns on an asset move with the market, there will be a component of risk (systematic risk) that cannot be diversified away. Assuming optimal portfolio diversification, the risk premium reflects the asset's systematic risk and hence its contribution to the overall variability of returns on the portfolio. This premium will be included in the equilibrium expected rate of return on this asset, in addition to the general rate of return on the portfolio.

The algebra of the CAPM theory is straightforward. All investors are assumed, as in standard portfolio analysis, to maximize their expected utility subject to their portfolio wealth constraint: this yields for each investor an optimally diversified holding of each asset in terms of the expected returns on all assets. It is convenient to aggregate this equation across investors and refer to the result as the asset demand of a 'typical' investor (but this does not imply that all investors are the same; it is merely an expression indicating averaging). The CAPM's contribution is then to note that at any time there will be a certain stock of each asset outstanding in the market which must be held in equilibrium. To be held, its expected return (and so its price) must satisfy the asset demand equation of the typical investor: in other words, this equation is turned round and solved for the necessary expected return in terms of the outstanding asset quantity. As the quantity of each asset rises, however risky it may be, it contributes only a little to the total risk on the whole market portfolio because it forms only a small part of this diversified whole; hence the rise in necessary expected return (fall in price) is negligible and we can talk of 'the' required expected return on an asset, independently of its quantity outstanding. Clearly this must not be pushed too far: some assets (e.g. dollar liabilities in currency

[2] Consider an individual faced with a possible gamble; he may choose either to receive 100 for sure, or to toss a coin and receive 50 if heads occur and 150 if tails occur. The expected outcome of this latter choice is 100 = 0.5(50) + 0.5(150). The question is: will the individual prefer the actuarial value of the gamble (this it its expected outcome) with certainty or will he prefer the gamble itself? If he prefers the gamble he is a risk lover; if he is indifferent he is risk neutral; and if he prefers the sure outcome he is risk averse. It is also possible to compute the maximum amount of wealth an individual would be willing to give up in order to avoid the gamble. This is the notion of a risk premium (see Pratt, 1964; Arrow, 1971; Markowitz, 1959).

portfolios) are outstanding in very large quantities and their risk cannot be fully diversified away.

Assume three assets, one of which yields a safe return . The typical investor maximizes the expected utility of end-of-period wealth, $E_tU(W_{t+1})$, with respect to w_1 and w_2 subject to the budget constraint

$$W_{t+1} = w_1R_{1t} + w_2R_{2t} + (1 - w_1 - w_2)R_t \tag{1}$$

where w_i is the share in the investor's portfolio of asset i, R_{it} is the actual return during the period, $W_{t+1}$is his end-period wealth expressed as an index (beginning period wealth = 1), U is his utility function ($U' > 0,\ U'' < 0$), and E_t is his expectation formed at the beginning of the period when he takes his investment decision.

Take a Taylor series expansion of $U(W_{t+1})$ around $E_t(W_{t+1})$:

$$U(W_{t+1}) = U(E_tW_{t+1}) + U'.(W_{t+1} - E_tW_{t+1}) + \\ 0.5U''.E_t(W_{t+1} - E_tW_{t+1})^2 + ... \tag{2}$$

Ignoring terms of higher order than two, (see e.g. Kraus and Litzenberger (1976), Hwang and Satchell (1997) or Harvey and Siddique (2000) for some implications of higher moments) or alternatively assuming that the agent's utility function only depends on the first two moments, the expectation of this is:

$$E_tU(W_{t+1}) = U(E_tW_{t+1}) + U'.(E_tW_{t+1} - E_tW_{t+1}) + \\ 0.5U''.E_t(W_{t+1} - E_tW_{t+1})^{"} \tag{3}$$

which is equal to

$$U(w_1E_tR_{1t} + w_2E_tR_{2t} + \{1 - w_1 - w_2\}R_t) + \\ 0.5U''(w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_{12}) \tag{4}$$

where $\sigma_{ij} = E_t([R_{it} - E_tR_{it}][R_{jt} - E_tR_{jt}])$, the covariance between the returns of assets i and j, and σ_i^2 = the variance of i's return.

The first order conditions with respect to w_1 and w_2 respectively are:

$$U'[E_tR_{1t} - R_t] = U''[w_1\sigma_1^2 + \sigma_{12}w_2] \tag{5}$$

and

$$U'[E_tR_{2t} - R_t] = U''[w_2\sigma_2^2 + \sigma_{12}w_1] \tag{6}$$

$$\frac{-U''}{U'} = p \tag{7}$$

p is defined as a measure of absolute risk aversion.

We can express (5) and (6) in the CAPM form with required expected asset returns as the dependent variable (the standard portfolio analysis has the asset shares as the dependent variables): w_1 and w_2 are given by the outstanding stocks of assets 1, 2 and the safe asset in the overall market.

Rearranging (5) the expected return on asset 1 for example is :

$$E_t R_{1t} = R_t + p(w_1 \sigma_1^2 + w_2 \sigma_{12}) = R_t + pw(x_1[\sigma_1^2 - \sigma_{12}] + \sigma_{12}) \quad (8)$$

and on asset 2

$$E_t R_{2t} = R_t + p(w_2 \sigma_2^2 + w_1 \sigma_{12}) = R_t + pw(x_2[\sigma_2^2 - \sigma_{12}] + \sigma_{12}) \quad (9)$$

where x_1 is the share of asset 1 in the risky part of the market portfolio, $x_1 = \frac{w_1}{w}$, $x_2 = \frac{w_2}{w}$, and w is the share of the risky part in the total market ($w = w_1 + w_2$). Suppose we regard asset 1 as being a single asset and asset 2 as being a portfolio of all assets in the risky market, that is essentially 'the' risky market portfolio. Then (8) reveals that the expected return on a single asset consists of the safe return plus a risk premium reflecting risk aversion (p), the overall share of risky assets in the whole market (w) and the covariance between asset 1 and the risky market (σ_{12}), 'systematic risk': there is also a small component in the risk premium for the extent to which risk on asset 1 exceeds this covariance ('diversifiable risk'), but this has a negligible effect because it is multiplied by x_1, the small share of asset 1 in the risky market. So provided asset 1 is small in relation to the market, its risk can be virtually totally diversified away and barely affects the expected return.

Suppose we now let asset 2 embrace the whole risky market. Then $x_2 = 1$ and we have from (9):

$$E_t R_{mt} = R_t + pw\sigma_m^2 \quad (10)$$

where m is the whole risky portfolio so that in this case $\sigma_2^2 = \sigma_m^2$. From this it follows that:

$$pw = \frac{(E_t R_{mt} - R_t)}{\sigma_m^2} \quad (11)$$

This means we can rewrite asset 1's required return, substituting (10) into (8), as

$$E_t R_{1t} = R_t + \frac{\{x_1[\sigma_1^2 - \sigma_{1m}] + \sigma_{1m}\}}{\sigma_m^2}(E_t R_{mt} - R_t) \quad (12)$$

or more generally for any asset

$$E_t R_{it} = R_t + \beta_i (E_t R_{mt} - R_t) \quad (13)$$

where we assume the weight x_i is small enough to be ignored and so

$$\beta_i = \frac{\sigma_{im}}{\sigma_m^2} \tag{14}$$

$E_t R_{mt} - R_t$ is the excess return required on a unit of the average portfolio: it is therefore 'the cost of average risk'. Individual assets command higher or lower excess returns according to their 'beta', β_i, which measures their systematic risk, the covariance of the ith asset with the market portfolio, divided by the variance of the market portfolio — as illustrated in figure 14.1.

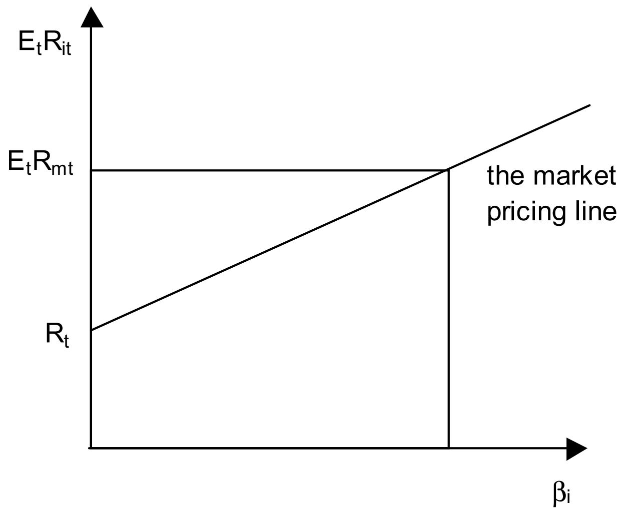

Figure 14.1: The Pricing of a Security

OTHER MODELS OF EQUILIBRIUM EXPECTED RETURNS

A related idea to the CAPM assumes that aggregate market risk is dominated by several sources of risk: principal component analysis can be used to separate these out. Thus there is not one but several market 'risks' or variables which are uncorrelated (for example, one variable might be a measure of the world business cycle, another might be the North-South terms of trade): the expected return associated with taking each risk is priced exactly as above — now think of 'asset 1' and 'asset 2' as being respectively 'exposure' to 'risk 1' and 'risk 2': that is, you buy £1 worth of assets combined in such a way that their return is correlated perfectly with variable 1, and similarly for variable 2. These fundamental risks being priced, each asset is then priced, by arbitrage, according to its amount of each risk. This is the arbitrage pricing theory (APT), due to Ross (1976).

CONSUMPTION CAPM

In chapters 11 and 12 we set out Lucas' theory of asset pricing. The first-order condition for the optimal consumption and portfolio decision is given by equation (9) in chapter 11 as

$$u'(c_t) = E_t\beta(1 + R_{it})u'(c_{t+1}) \tag{15}$$

where we have generalised the model to allow for the real gross return, $1 + R_{it}$, on the ith asset'[3]

Recalling that the covariance of X and Y denoted by $covXY$ is by definition $covXY \equiv E(X - EX).(Y - EY) \equiv EXY - EX.EY$ (we also note in passing that $cov(1 + X,\ Y) = covXY$), we can write the unconditional form of (15) as

$$E(1 + R_{it}) = \frac{(1 - cov(1 + R_{it}.K_{t+1}))}{EK_{t+1}} \tag{16}$$

[3] If we write the budget constraint as
$c_t + p_t s_{t+1} + q_t b_{t+1} = (p_t + d_t)s_t + b_t$
where b_t represents holdings of the riskless asset, s_t represents holdings of the risky asset, p_t is the price of the risky asset, q_t is the normalised price of the riskless asset. The optimization yields

$$u'(c_t) - \lambda_t = 0 \tag{1}$$
$$-\lambda_t q_t + \beta E_t \lambda_{t+1} = 0 \tag{2}$$
$$-\lambda_t p_t + \beta E_t(p_{t+1} + d_{t+1})\lambda_{t+1} = 0 \tag{3}$$

where λ_t is the Lagrange multiplier.
These conditions, 1-3, yield

$$\begin{aligned} 1 &= \beta E_t \frac{u'(c_{t+1})}{q_t u'(c_t)} \\ 1 &= \beta E_t \left[\frac{(p_{t+1} + d_{t+1})u'(c_{t+1})}{p_t u'(c_t)}\right] \end{aligned}$$

noting that $q_t^{-1} = (1 + R)$ where R is the risk-free real rate. p, s can be generalized to any asset as in the text.

where $K_{t+1} = \frac{\beta u'(c_{t+1})}{u'(c_t)}$ is called the stochastic discount factor or pricing kernel.

For a risk free real interest rate R_t^f the covariance term is zero so that

$$E(1+R_t^f) = \frac{1}{EK_{t+1}} \tag{17}$$

For the portfolio of assets we have that

$$E(1+R_{mt}) = \frac{(1 - cov(R_{mt}, K_{t+1}))}{EK_t} \tag{18}$$

We can employ (16), (17) and (18) to obtain

$$E(R_{it}) - R_t^f = -(1+R_t^f)cov(R_{it}, K_{t+1}) \tag{19}$$

and

$$E(R_{it}) - R_t^f = \frac{[covR_{it}, K_{t+1}](ER_{mt} - R_t^f)}{cov(R_{mt}, K_{t+1})} \tag{20}$$

Equation (20) illustrates that an asset's expected real return in excess of the risk free rate is higher the more negative the covariance between the asset and the ratio of marginal utilities. This is because when consumption is smaller, marginal utility is higher. An asset with a more negative covariance thus implies that the return is lower when consumption is low. This is the precisely the state however when wealth is more valued; consequently the agent requires a higher risk premium to hold the asset. This model of the relationship between expected returns is known as the consumption CAPM. Equation (20) shows the relationship to the ordinary CAPM where covariances with the pricing kernel replace those with the market portfolio.

To make the model empirically operational we have to specify the error structure of the model and the form of the utility function.

Consider the implications of the model if

$$u(c_t) = \frac{c_t^{1-\delta} - 1}{1-\delta} \tag{21}$$

where δ is a constant and is called the coefficient of relative risk aversion.[4] This function has the convenient property that it nests the linear case, $\delta = 0$, and the logarithmic form, $\delta = 1$. (The latter case is found by employing L'Hopital's rule when $\delta = 1$.)

For this function

$$\frac{u'(c_{t+1})}{u'(c_t)} = (\frac{c_{t+1}}{c_t})^{-\delta} \tag{22}$$

In addition we assume that consumption and asset returns are jointly conditionally lognormally distributed. This allows us to make use of the fact that for a conditionally lognormally distributed variable, Y,

$$\ln E_t(Y) = E_t[\ln Y] + 0.5var_t[\ln Y] \tag{23}$$

Taking logarithms of (15) and employing (23) we obtain

$$\begin{aligned} E_t \ln(1+R_{it}) = &-\ln\beta + \delta E_t(\ln c_{t+1} - \ln c_t) \\ &- 0.5\{\text{var}\ln(1+R_{it}) + \delta^2 \text{var}(\ln \frac{c_{t+1}}{c_t})) \\ &- 2\delta \text{cov}(\ln(1+R_{it})\ln(\frac{c_{t+1}}{c_t}))\} \end{aligned} \tag{24}$$

(Note let $Y_t = \frac{(1+R_{it})u'(c_{t+1})}{u'(c_t)}$ so from (15), $\ln(1) = \ln\beta + \ln(E_t Y_t)$. Then use (22) and (23) to obtain

$$\begin{aligned} 0 \;=\; &\ln\beta + E_t(\ln(1+R_{it}) - \delta\ln(\frac{c_{t+1}}{c_t})) + 0.5\{\text{var}(1+R_{it}) + \\ &\delta^2 \text{var}(\frac{c_{t+1}}{c_t}) - 2\delta \text{cov}(\ln(1+R_{it})\ln(\frac{c_{t+1}}{c_t}))\} \end{aligned}$$

Rearranging this expression gives (25))

If the real rate of interest on an asset, R_t^f, is risk free (24) simplifies to

$$\ln(1+R_t^f) = -\ln\beta + \delta E_t(\ln c_{t+1} - \ln c_t) - 0.5(\delta^2 \text{var}(\ln \frac{c_{t+1}}{c_t}) \tag{25}$$

When there is a nominal asset yielding a certain nominal rate, N_t, we can employ the definition of the real rate of interest for this asset, say, $1 + R_{Nt}$, as

$$1 + R_{Nt} \equiv \frac{(1+N_t)P_t}{P_{t+1}} \tag{26}$$

where P is the price level.

Substitution of (26) in (15) and employing the same method as above gives us the required nominal return on the safe nominal asset, assuming conditional normality, as:

$$\begin{aligned} \ln(1+N_t) = &-\ln\beta + E_t(\ln P_{t+1} - \ln P_t) + \delta E_t(\ln c_{t+1} - \ln c_t) \\ &- 0.5(\delta^2 \text{var}(\ln \frac{c_{t+1}}{c_t}) + \text{var}(\pi) + 2\delta cov(\ln \frac{c_{t+1}}{c_t}, \pi)) \end{aligned} \tag{27}$$

where $\pi = \ln(P_{t+1}) - \ln(P_t)$

We observe from (25) that the riskless real rate has a linear relationship with the expected change in real consumption. If the riskless real rate is assumed constant we can rearrange (25) with expected changes in consumption as the left hand side variable. This is the relationship estimated by Hall (1978): under rational expectations changes in consumption should be orthogonal to information dated prior to the information set conditioning expectations.

From (27) we observe that the safe nominal asset responds with a unit coefficient to expected inflation and is also determined by terms which reflect inflation risk premia.

Subtracting (25) from (24) we obtain

$$E_t \ln(1+R_{it}) - \ln(1+R_t^f) + 0.5 var(1+R_{it}) = \delta cov\left(\ln(1+R_{it}), \ln(\frac{c_{t+1}}{c_t})\right) \quad (28)$$

or that

$$\ln E_t \frac{(1+R_{it})}{(1+R_t^f)} = \delta cov\left(\ln(1+R_{it}), \ln(\frac{c_{t+1}}{c_t})\right) \quad (29)$$

The implication of this model is that risk premia are determined by the covariation between the asset and changes in consumption times the coefficient of relative risk aversion.

Because the covariance term is relatively small in actual data, empirical tests of this model have not been able to explain the excess return on stocks over bonds, on average some 6% in the US over 100 years and large in other countries as well (see Campbell, 1996), without assuming a possibly implausibly large coefficient of relative risk aversion (well in excess of 10). This has become known as the equity premium puzzle after the pioneering contribution of Mehra and Prescott (1985). Recent contributions have suggested a variety of mechanisms that might resolve the puzzle. These include more general specifications of the utility function (e.g. Epstein and Zin, 1989, 1991), the introduction of habit formation (e.g. Constantinides, 1990; Campbell and Cochrane, 1999, 2000), survivor bias (e.g. Brown, Goetzmann and Ross, 1995), peso problems (see below; Rietz, 1988), heterogeneous agents (e.g. Constantinides and Duffie, 1996). Many of these ideas are superbly set out in Campbell, Lo and MacKinlay (1997).

DETERMINATION OF ASSET PRICE BEHAVIOUR UNDER RATIONAL EXPECTATIONS

We next consider the implications of the assumption of rational expectations for the behaviour of asset prices. Consider the return to holding an asset over the period t to $t+1$, say a share. We have

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} - 1 \tag{30}$$

where D_{t+1} is any dividend or payment in the period and $P_t,\ P_{t+1}$ is the price of the asset at time t and $t+1$.

We assume initially for simplicity that investors are risk neutral so that via arbitrage expected returns are equal to those on a riskless asset with a real rate of interest, $\overline{R}$, assumed constant, so that $E_t R_{t+1} = \overline{R}$.

Since P_t is part of the current information set, the expectation of (30) when rearranged is given by

$$P_t = \frac{1}{(1+\overline{R})} E_t P_{t+1} + \frac{1}{(1+\overline{R})} E_t D_{t+1} \tag{31}$$

Assuming rational expectations and solving this model forwards N periods we obtain

$$P_t = E_t \sum_{i=1}^{N} \frac{D_{t+i}}{(1+\overline{R})^i} + \frac{E_t P_{t+N}}{(1+\overline{R})^N} \tag{32}$$

In the absence of speculative bubbles (see below) we assume that the last term goes to zero as we let N go to infinity. In this case the current asset price is equal to the expected value of the stream of dividends into the indefinite future. This term is the fundamental of the process which we can call F_t.

Under these assumptions we have that:

$$P_t = E_t \left[\sum_{i=1}^{\infty} \frac{D_{t+i}}{(1+\overline{R})^i} \right] \tag{33}$$

We can write (33) in the equivalent form

$$P_t = E_t \left[\sum_{i=1}^{\infty} \frac{D_{t+i}}{(1+\overline{R})^i} \right] + \frac{D_t}{1+\overline{R}} - \frac{D_t}{1+\overline{R}} + \frac{E_t D_{t+1}}{(1+\overline{R})^2} - \frac{E_t D_{t+1}}{(1+\overline{R})^2} + \frac{E_t D_{t+2}}{(1+\overline{R})^3} - \frac{E_t D_{t+2}}{(1+\overline{R})^3} + \ldots \tag{34}$$

Rearranging (34) we have that

$$P_t = E_t\left[\sum_{i=1}^{\infty}\frac{\Delta D_{t+i}}{(1+\overline{R})^i}\right] + \frac{D_t}{1+\overline{R}} + \frac{E_t D_{t+1}}{(1+\overline{R})^2} + \frac{E_t D_{t+2}}{(1+\overline{R})^3} + \dots \quad (35)$$

so that using (33) we can rewrite (35) as

$$P_t = E_t\left[\sum_{i=1}^{\infty}\frac{\Delta D_{t+i}}{(1+\overline{R})^i}\right] + \frac{D_t}{1+\overline{R}} + \frac{P_t}{1+\overline{R}} \quad (36)$$

Multiplying through by $1+\overline{R}$ and rearranging we obtain the alternative form

$$P_t = \frac{D_t}{\overline{R}} + \frac{1}{\overline{R}}E_t\left[\sum_{i=1}^{\infty}\frac{\Delta D_{t+i}}{(1+\overline{R})^{i-1}}\right] \quad (37)$$

Equation (37) shows that the current price of the stock is equal to the dividend divided by $\overline{R}$ plus a term in the discounted stream of expected future changes in dividends. In this form the model is amenable to empirical testing in the form of cointegration analysis (see Time-Series Annex). If dividends are a non-stationary process but changes in dividends are stationary then the stock price is cointegrated with dividends with coefficient $\frac{1}{\overline{R}}$.

An insightful special case of the above model arises when dividends are expected to grow at a constant rate g.

For this case

$$E_t D_{t+i} = (1+g)E_t D_{t+i-1} = (1+g)^i D_t \quad (38)$$

Substituting (38) in (33) we obtain

$$P_t = \frac{(1+g)D_t}{(1+\overline{R})} + \frac{(1+g)^2 D_t}{(1+\overline{R})^2} + \frac{(1+g)^3 D_t}{(1+\overline{R})^3} + \frac{(1+g)^4 D_t}{(1+\overline{R})^4} + \dots \quad (39)$$

We can rewrite (39) as

$$P_t = \frac{(1+g)D_t}{(1+\overline{R})}\left[1 + \frac{(1+g)}{(1+\overline{R})} + \frac{(1+g)^2}{(1+\overline{R})^2} + \frac{(1+g)^3}{(1+\overline{R})^3} + \dots\right] \quad (40)$$

Recalling that $\frac{1}{1-x} = 1 + x + x^2 + x^3 + ...$ for $\mid x \mid < 1$ and assuming that $g < \overline{R}$ (as it must be since the stock price is not infinite) we can rewrite (40) as

$$P_t = \frac{(1+g)D_t}{(1+\bar{R})}\left\{\frac{1}{1-\frac{(1+g)}{(1+\bar{R})}}\right\} = \frac{(1+g)D_t}{\overline{R}-g} \quad (41)$$

This form is the Gordon growth model. It demonstrates the important point that if $\bar{R}$ is close to g then small permanent changes in $\overline{R}$ can have a major impact on the stock price.

The assumption made in the derivation of the asset price (33) is that expected stock returns are equal to a constant risk-free real rate of interest (in the context of CAPM say a constant real rate plus a constant risk premium). Relaxing this assumption in the above framework results in the loss of analytic tractability since expectations would be of a non-linear form. It is interesting to note an approximation for the case of a variable return that preserves tractability which was initially employed by Campbell and Shiller (1988 a, b).

Taking logarithms of (30) we have that

$$\ln(1+R_{t+1}) = \ln(P_{t+1}+D_{t+1}) - \ln(P_t) = \ln(P_{t+1}) - \ln(P_t) + \ln\left(1+\frac{D_{t+1}}{P_{t+1}}\right) \quad (42)$$

Noting that $\frac{D_{t+1}}{P_{t+1}} = \exp[\ln D_{t+1} - \ln P_{t+1}]$ we can rewrite (42) as

$$r_{t+1} = p_{t+1} - p_t + \ln(1+\exp(d_{t+1}-p_{t+1})) \quad (43)$$

where lower case letters now represent logarithms so that $r_t = \ln(1+R_t)$

Letting $z_t = d_{t+1} - p_{t+1}$ and taking a first-order Taylor expansion of the last term in (43) around the mean value of $\overline{z}$ we obtain

$$r_{t+1} = p_{t+1} - p_t + \ln(1+\exp(\overline{z})) + \frac{\exp(\overline{z})}{1+\exp(\overline{z})}[d_{t+1} - p_{t+1} - \overline{z}] \quad (44)$$

Rearranging (44) we obtain

$$r_{t+1} = \frac{p_{t+1}}{1+\exp(\overline{z})} + \frac{\exp(\overline{z})d_{t+1}}{1+\exp(\overline{z})} - p_t + \ln(1+\exp(\overline{z})) - \frac{\overline{z}\exp(\overline{z})}{1+\exp(\overline{z})} \quad (45)$$

Letting $\lambda = \frac{1}{1+\exp(\overline{z})}$ and noting that $\frac{1-\lambda}{\lambda} = \exp(\overline{z})$ and $1-\lambda = \frac{\exp(\overline{z})}{(1+\exp(\overline{z}))}$ we can rewrite (45) as

$$r_{t+1} = \lambda p_{t+1} + (1-\lambda)d_{t+1} - p_t - \ln(\lambda) - (1-\lambda)\ln\left(\frac{1-\lambda}{\lambda}\right) \quad (46)$$

or

$$r_{t+1} = \lambda p_{t+1} + (1-\lambda)d_{t+1} - p_t + \delta \quad (47)$$

where $\delta = -\ln(\lambda) - (1-\lambda)\ln\left(\frac{1-\lambda}{\lambda}\right)$

For a constant dividend-price ratio

$$\lambda = \frac{1}{1 + \frac{D}{P}} \tag{48}$$

The average value of the dividend price ratio observed in economies shows that the value of λ will be close to unity, around 0.96 in the US for example, so that in (47) the weight on the log price is close to one and that on log dividends closer to zero. Clearly the approximation will be more accurate the smaller the variation in the log dividend price ratio. Analysis by Campbell and Shiller (1988a) is suggestive that the approximation does not produce gross violations of reality, particularly at the monthly level of analysis. Essentially the approximation provides analytical tractability at the cost of some error in the statement of average returns. From an empirical perspective dividends appear to be more parsimoniously explained as a loglinear rather than a linear time-series process and this is an advantage of the approximation method.

We can take expectations of (47) and solve forward to obtain

$$p_t = \frac{\delta}{1-\lambda} + E_t \left[\sum_{i=0}^{\infty} \lambda^i \left[(1-\lambda)\, d_{t+i+1} - r_{t+i+1} \right] \right] \tag{49}$$

where we assume once again in the absence of speculative bubbles that the term $E_t \lambda^{\infty} p_{t+\infty}$ is zero.

We observe from (49) that the current log stock price is higher the higher expected future dividends and the lower expected future returns (i.e. the stock discount factors).

We can transpose (49) following the type of procedure outlined above (34) to obtain the equivalent form

$$p_t = \frac{\delta}{1-\lambda} + d_t + E_t \left[\sum_{i=0}^{\infty} \lambda^i \left(\Delta d_{t+i+1} - r_{t+i+1} \right) \right] \tag{50}$$

(Hint: in (49) add inside square brackets the term $d_t - d_t + \lambda d_{t+1} - \lambda d_{t+1} + \lambda^2 d_{t+2} - \lambda^2 d_{t+2} + - + \lambda^n d_{t+n} - \lambda^n d_{t+n}$, rearrange in terms of Δd_{t+1+i}. The terms in $d_t + \lambda d_{t+1} + \lambda^2 d_{t+2} + ... + \lambda^n d_{t+n}$then can be substituted out.)

Equation (47) is useful for illustrating another point.Taking expectations at time t of (47) and subtracting the resultant from (47) we obtain

$$r_{t+1} - E_t r_{t+1} = \lambda (p_{t+1} - E_t p_{t+1}) + (1-\lambda)[d_{t+1} - E_t d_{t+1}] \tag{51}$$

If we lead (50) one period, take expectations of it at time t and subtract the resultant from it, we obtain an expression for $p_{t+1} - E_t p_{t+1}$ which we substitute in (51) to obtain

$$r_{t+1} - E_t r_{t+1} =$$
$$\lambda \left\{ \begin{array}{c} d_{t+1} - E_t d_{t+1} + E_{t+1} \left[\sum_{i=0}^{\infty} \lambda^i \left(\Delta d_{t+2+i} - r_{t+2+i} \right) \right] \\ -E_t \left[\sum_{i=0}^{\infty} \lambda^i \left(\Delta d_{t+2+i} - r_{t+2+i} \right) \right] \end{array} \right\}$$
$$+ (1 - \lambda)\left(d_{t+1} - E_t d_{t+1} \right) \quad (52)$$

Recognising that $d_{t+1} - E_t d_{t+1} = E_{t+1} d_{t+1} - E_{t+1} d_t - (E_t d_{t+1} - E_t d_t)$ under the information assumptions of the model, i.e. d_t is observable at time t, d_{t+1} at $t+1$, so $E_{t+1} d_{t+1} = d_{t+1}$, $E_t d_t = d_t$, we can rewrite (52) in the form presented by Campbell (1991) namely

$$r_{t+1} - E_t r_{t+1} = E_{t+1} \sum_{i=0}^{\infty} \lambda^i \Delta d_{t+1+i} - E_t \sum_{i=0}^{\infty} \lambda^i \Delta d_{t+1+i}$$
$$- [E_{t+1} \sum_{i=1}^{\infty} \lambda^i r_{t+1+i} - E_t \sum_{i=1}^{\infty} \lambda^i r_{t+1+i}]$$
$$= (E_{t+1} - E_t) \sum_{i=0}^{\infty} \lambda^i (\Delta d_{t+1+i} - \lambda r_{t+2+i}) \quad (53)$$

Equation (53) is informative. We observe that unexpected stock returns are a function of revisions of expected future changes in dividends and revisions of stock discount factors. Further, *ceteris paribus* an upward revision in expectations of future returns leads to a fall in the price today since for a given dividend stream this can only be generated from a lower price today. The equation also demonstrates how revisions of expectations into the indefinite future impact on the current innovation in returns. Under rational expectations these revisions are news. However this creates a problem in empirical work which endeavours to estimate the impact of empirical measures of news on unanticipated asset returns. Clearly the impact of current news can have an ambiguous impact on unanticipated returns depending on the implications it has for future revisions to expectations. For example the news that the current money stock is higher than anticipated could be interpreted as a signal for future tightening of money stock or a move to a more relaxed regime. Similarly output figures higher than anticipated could signal the end of a slump or the beginning of an inflationary episode. This implies that without knowledge of current and future policy regimes empirical estimates of the impact of current news will be 'non-structural' so that the estimated coefficients may be difficult to interpret.

We can illustrate two other important implications for the potential behaviour of asset prices and returns using a modified example borrowed from Campbell, Lo and Mackinlay (1997).

Assume that the log of dividends is generated by the process

$$d_{t+1} = \rho d_t + u_{t+1} \tag{54}$$

where u_{t+1} is a random variable and$\mid \rho \mid < 1$. So we assume that dividends are a stationary process in this example.

Further assume that the stock discount factor is generated by

$$E_t r_{t+1} = \overline{r} + y_t \tag{55}$$

where $\overline{r}$ is a constant and y_t is generated by the process

$$y_t = \alpha y_{t-1} + v_t \tag{56}$$

where $\mid \alpha \mid < 1$ and v_t is a random variable.

y_t is an observable variable and might represent changing risk. Given these assumptions we can substitute in (50) to obtain the solution for the price of the asset. The log price of the asset (50) is given by

$$p_t = \frac{\delta}{1-\lambda} + E_t \left\{ \sum_{i=0}^{\infty} \lambda^i \left[(1-\lambda) d_{t+1+i} - r_{t+1+i} \right] \right\} \tag{57}$$

The first term in the braces, substituting for future dividends, is given by

$$\begin{aligned} &(1-\lambda)[E_t d_{t+1} + \lambda E_t d_{t+2} + \lambda^2 E_t d_{t+3} + \ldots] = \\ &\qquad\qquad (1-\lambda)[\rho d_t + \lambda \rho^2 d_t + \lambda^2 \rho^3 d_t + \ldots] \end{aligned} \tag{58}$$

and the second term by

$$\begin{aligned} &-\{E_t r_{t+1} + \lambda E_t r_{t+2} + \lambda^2 E_t r_{t+3} + \ldots\} = \\ &\qquad\qquad -\{\overline{r} + y_t + \lambda(\overline{r} + \alpha y_t) + \lambda^2(\overline{r} + \alpha^2 y_t) + \ldots\} \end{aligned} \tag{59}$$

so that substituting in (57) we obtain

$$p_t = \frac{\delta}{1-\lambda} + d_t \frac{(1-\lambda)\rho}{1-\lambda\rho} - \frac{\overline{r}}{1-\lambda} - \frac{y_t}{1-\lambda\alpha} \tag{60}$$

(recalling again that $\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots$ for $\mid x \mid < 1$)

We can substitute (60) into (47) to obtain the solution for the one period stock return as

$$\begin{aligned} r_{t+1} = {} & \lambda \left(\frac{\delta}{1-\lambda} + d_{t+1} \frac{(1-\lambda)\rho}{1-\lambda\rho} - \frac{\overline{r}}{1-\lambda} - \frac{y_{t+1}}{1-\lambda\alpha} \right) + (1-\lambda) d_{t+1} \\ & - \left(\frac{\delta}{1-\lambda} + d_t \frac{(1-\lambda)\rho}{1-\lambda\rho} - \frac{\overline{r}}{1-\lambda} - \frac{y_t}{1-\lambda\alpha} \right) + \delta \end{aligned} \tag{61}$$

Using (54) and (56) and simplifying this gives

$$r_{t+1} = \overline{r} + u_{t+1} + y_t - \frac{\lambda v_{t+1}}{(1-\lambda\alpha)} \tag{62}$$

The solution for log prices illustrates that when the variability of y_t is small, so that the variability of expected returns is small, this does not imply that the variability of prices need be small. This is because the denominator of the term in y_t can be small. More formally, assuming the covariance between y_t and d_t is zero, we have from (60) that the variance of prices is given by

$$\sigma_p^2 = \frac{\sigma_y^2}{(1-\lambda\alpha)^2} + \frac{(1-\lambda)^2\rho^2}{(1-\lambda\rho)^2}\sigma_d^2 \text{ where } \sigma_y^2 = \frac{\sigma_v^2}{1-\alpha^2} \tag{63}$$

where $\sigma_d^2 = \frac{\sigma_u^2}{1-\rho^2}$ is the variance of dividends.

Clearly σ_y^2 can be small but σ_p^2 large.

We can also solve for the reduced form ARMA time-series representation (see Time-Series Annex) of log returns and log prices in our example.

Using the lag operator, and substituting for y_t we obtain for returns

$$r_{t+1} = \overline{r} + u_{t+1} + \frac{v_t}{(1-\alpha L)} - \frac{\lambda v_{t+1}}{(1-\lambda\alpha)} \tag{64}$$

Multiplying out the lag operator we obtain

$$r_{t+1} = \alpha r_t + \overline{r}(1-\alpha) + u_{t+1} - \alpha u_t + \frac{v_t}{(1-\lambda\alpha)} - \frac{\lambda v_{t+1}}{(1-\lambda\alpha)} \tag{65}$$

The summation of the two error terms on the right hand side of (65) can be rewritten as a moving average error process of order one so that observed returns follow an ARMA (1, 1) process. Depending upon the covariance between news in dividends and news in returns the process can exhibit positive or negative autocorrelation.

Employing a similar process for changes in asset prices we obtain the ARMA representation for the level of the log asset price as (where L is the lag operator)

$$p_t(1-\rho L)(1-\alpha L) = \frac{(\delta-\overline{r})(1-\alpha)(1-\rho)}{(1-\lambda)} + \frac{(1-\lambda)\rho u_t(1-\alpha L)}{1-\lambda\rho}u_t - \frac{v_t(1-\rho L)}{(1-\lambda\alpha)} \tag{66}$$

so that the level of the log of the asset price is described by an ARMA(2, 1) in this example.

Although both returns and asset prices are forecastable in this example there can be no presumption that the market is inefficient. The crucial element in the definition is that *abnormal* returns should not be forecastable in an efficient market. We will observe another example of this next where the exchange rate has a predictable path in the Dornbusch overshooting model even though the market is efficient by construction.

OPEN ECONOMY MODELS WITH EFFICIENT FINANCIAL MARKETS

In Chapter 10 we set out the general behaviour of macro models of the open economy with efficient markets and New Classical price behaviour, under fixed and floating exchange rates. Here we focus in more detail on the behaviour of nominal exchange rates under floating, under varying assumptions about price behaviour. This will illustrate the role of financial efficiency per se in open macro models.

Our model is based on that outlined by Dornbusch (1976) in his seminal paper. For simplicity it is assumed that there is perfect capital mobility between countries (i.e. transactions costs are negligible and international assets are perfect substitutes). Consider initially a risk-neutral agent who is faced with the choice between holding a domestic or foreign bond for the duration of one period (say 90 days). The nominal rates of interest in the foreign country and in the domestic country are given by R_t and R_t^F respectively. Since the bonds are perfect substitutes, asset market equilibrium requires that the expected rates of return on the two bonds be equal. This expected rate of return has two components. The first component is the interest rate on the bond, which we can assume to be known at t; the second component is the expected capital gain or loss from exchange rate changes during the 90-day period.

It follows that the speculative condition for equilibrium, known as uncovered interest arbitrage is:

$$R_t = R_t^F + (E_t S_{t+1} - S_t) \tag{67}$$

where R_t, R_t^F are the domestic and foreign nominal interest rate, S_t is the logarithm of the current exchange rate (here domestic units per foreign unit) and $E_t S_{t+1}$is the expectation of the rate in period one (90 days in our example). A rise in S_t in our notation here represents a depreciation of the home currency. Equation (67) therefore implies that the interest rate differential in favour of domestic bonds must be equal to the expected depreciation of the exchange rate.

For example, if the domestic currency pays 12 per cent interest (per 90 days) and the foreign currency pays 4 per cent interest, a domestic investor buying foreign currency at the beginning of the period and converting back at the end of the period will, assuming the domestic currency depreciates by 8 per cent, expect to finish up with sufficient domestic currency to make him indifferent between holding domestic or foreign bonds.

There is also a forward market for foreign exchange in many exchange rates (i.e. traders can at time t contract to trade foreign currency at time $t+1$). In the forward market, large transactors are required to put up only very small amounts of money as 'margin requirements', so there is no need to discount. Consequently, using a similar argument to the one above, arbitrage implies the covered interest arbitrage condition that

$$R_t = R_t^F + (F_t - S_t) \tag{68}$$

where F_t is the logarithm of the forward exchange rate at time t for period $t+1$. We note that the covered condition is riskless and holds via arbitrage regardless of the manner in which expectations are formed. Uncovered arbitrage is a speculative condition, hence the explicit assumption of risk-neutral investors. Equating the covered and uncovered condition we obtain that $F_t = E_t S_{t+1}$ so that the forward rate is a direct measure of the market's expectation of the future exchange rate. The properties of the forward rate as a predictor of future spot rates has been a focus of much empirical research as we will discuss below.

In the Dornbusch model it is assumed that prices in goods or labour markets are in the short term 'sticky' with respect to changes in market conditions. This could, for instance, be because of the existence of multi-period wage or price contracts as in the New Keynesian model. It follows from this assumption that purchasing power parity does not hold in the short run. Purchasing power parity (PPP) or the 'law of one price', states that in the absence of transport costs and other transactions costs international arbitrage in goods should eliminate differentials between the prices of goods in different countries. We discussed in chapter 10 on the open economy how this would not hold in the short run but should hold in the long run, whether traded goods are homogeneous across countries or differentiated and imperfectly competitive (PPP in the long run but not the short is confirmed by numerous empirical tests, e.g. Taylor et al., 2001).

Under PPP we would have:

$$S_t = p_t - p_t^F \tag{69}$$

where p, p^F are the logarithms of the domestic and foreign price level.

With perfect capital mobility and PPP, from (67) and (69), the equalization of expected real rates of interest immediately follows:

$$R_t - (E_t p_{t+1} - p_t) = R_t^F - (E_t p_{t+1}^F - p_t^F) \tag{70}$$

Absence of PPP in the short run means that we may examine the behaviour of the real exchange rate, x, defined as:

$$x_t = S_t - p_t + p_t^F \tag{71}$$

For simplicity we assume that output (in logs), real interest rates, foreign interest rates and the foreign price level (in logs) are fixed and normalised to zero; but this does not affect the features of the model on which we focus here. The model is given by

$$R_t = E_t S_{t+1} - S_t \tag{72}$$

$$m_t^s = m_t + \overline{m} = p_t - \delta R_t \tag{73}$$

$$p_t - p_{t-1} = k(S_t - p_{t-1}) + u_t \tag{74}$$

where m_t is a random monetary shock around $\overline{m}$, the constant average money supply, and u_t is a supply shock. Equation (73) is a conventional demand for money function equated to money supply. Equation (74) is an *ad hoc* price adjustment mechanism which captures the hypothesis that the price level responds sluggishly to changes in the exchange rate, via aggregate demand and its effects on the volume of net trade. The specification has the convenient property that when $k = 1$, PPP holds instantaneously apart from the current supply shock.

When PPP holds instantaneously the model can be rearranged to give

$$S_t = \frac{\delta E_t S_{t+1}}{(1+\delta)} + \frac{m_t^s - u_t}{(1+\delta)} \tag{75}$$

Equation (75) will be recognised to have the same form as (31) above. We solve this model under the assumption that agents have full current information. The reduced form for the model is given by

$$k\overline{m} + m_t - (1-k)m_{t-1} = kS_t + u_t - \delta(E_t S_{t+1} - S_t) + \delta(1-k)(E_{t-1}S_t - S_{t-1}) \tag{76}$$

Following the procedures outlined in chapter 2, the solution for the exchange rate in the full-information case is given by[4]

$$S_t = (1-z)\overline{m} + zS_{t-1} + a_o m_t + a_1 m_{t-1} + a_2 u_t \tag{77}$$

where z is the stable root of the equation:

$$\delta z^2 - z(\delta(1-k) + k + \delta) + \delta(1-k) = 0 \tag{78}$$

and $a_0 = \frac{1}{k+\delta(1-z)+\delta(1-k)},\ a_1 = -(1-k)a_0,\ a_2 = \frac{-1}{k+\delta(1-z)}$

We notice from (77) that, in the long run, when $S_t = S_{t-1}$, the elasticity of the exchange rate with respect to an increase in the permanent level of money supply is unity. In other words, in the long run the exchange rate depreciates by the change in $\overline{m}$. Given that $\overline{m}$ is a constant (72) implies that R is zero in the long run, and hence (73) that prices have the same response to $\overline{m}$ as the exchange rate. Equation (74) implies that in the long run PPP must hold. Leading (77) one period and taking expectations we obtain:

$$E_t S_{t+1} - S_t = (1-z)(\overline{m} - S_t) + a_1 m_t \tag{79}$$

Equation (77) defines an expectations mechanism known as regressive expectations. It informs us that when the equilibrium exchange rate is above the current exchange rate, then expectations are revised upwards and vice versa. The fact that regressive expectations can by choice of the regressive parameter $(1-z)$ be rational is one implication of this Dornbusch model. However, we should note that this property can hold in rational expectations models only where there is one stable root; it does not hold generally.[5]

We can substitute (77) into (71) and (73) to obtain

$$S_t = \frac{\overline{m}[1+\delta(1-z)] + m_t - u_t - (1-k)p_{t-1}}{k+\delta(1-z)} \tag{80}$$

---

[4] For the real interest rate differential to be constant, as investigated by Mishkin (1981), an infinitely large intertemporal substitution response is required for either α or β. The evidence does not support such responses; it is therefore not surprising that Mishkin concludes from his reduced-form work that the differential varies over time. Similar arguments apply to variation of the real interest rate within a closed economy, as investigated by Fama (1975) for the USA; this evidence has now been found (by Nelson and Schwert, 1977) to support non-constancy, which again is not surprising. We may also note that the size of variation in both the differential and the closed (e.g. world) economy level of real interest rates cannot be suggested a priori; nor can the length of time to convergence (determined by z in this model as influenced by all the parameters).

[5] Relative risk aversion (RRA) is defined as $RRA = \frac{-c_t u''(c_t)}{u'(c_t)} = \delta$ for the specification of the utility function.

Equation (80) illustrates another key insight of Dornbusch.

The impact of a change in the permanent level of money supply ($\overline{m}$) in the short run, ceteris paribus, is greater than unity except when $k = 1$ and consequently greater than the long-run impact. This phenomenon is known as 'overshooting'. The rationale for this effect is that because, in the short run, prices are at a point in time sticky or adjusting slowly, the only way the money market can remain in equilibrium as the permanent level of the money supply is increased is for the interest rate to fall. However, a falling interest rate has to be associated with an expected appreciation of the currency. Consequently the current exchange rate has to depreciate further than its long-run value in order to give rise to anticipations of an appreciation as it moves to its ultimate long-run value.

This mechanism is illustrated diagrammatically in figure 14.2. Suppose that at time $t = 0$ the pound/dollar rate is \$1 = £1. At time $t = N$ the authorities increase the level of the money supply by 100 per cent. In the long run this causes the pound to depreciate against the dollar to \$1 =£2.0. However, in the short run the pound depreciates further, to say \$1 = £2.50 and then follows the arrowed path back to long-run equilibrium. We notice that along the arrowed path the pound is appreciating, but has always depreciated relative to $t = 0$. The possibility that efficient assets markets, in conjunction with sticky wages or prices, could give rise to volatile behaviour of asset prices was a principal insight of Dornbusch and has been influential (see e.g. Buiter and Miller, 1981).

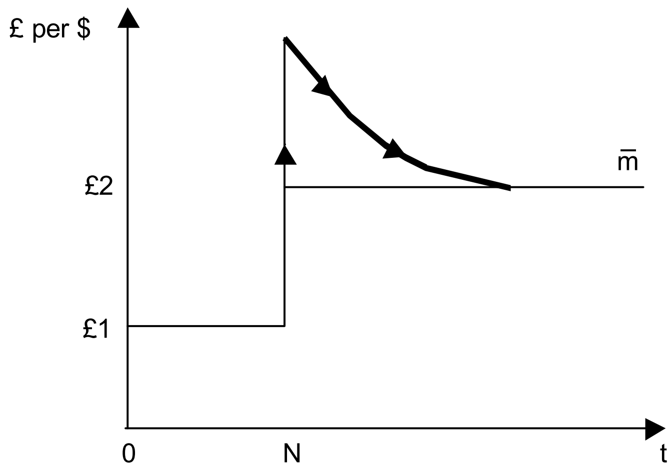

Figure 14.2: Exchange Rate Overshooting

We should also note from the solution for the exchange rate that, while a positive monetary shock causes a depreciation of the currency, its impact can be greater or less than unity, but since $\overline{m}$ has not changed, there will always be an 'overshoot' of the long-run equilibrium. The Dornbusch result applies to permanent (unanticipated) changes in the money stock ($\overline{m}$).

Another important feature of the model is that the exchange rate has a predictable path (77). However uncovered arbitrage was assumed in the model consequently the asset market is efficient since expected abnormal returns (in this example deviations from uncovered interest arbitrage) are not predictable. This is another example where predictability of an asset price does not violate market efficiency. In particular the exchange rate does not follow a random walk, though this is sometimes assumed in empirical work.

A number of authors have attempted to test the Dornbusch model (see e.g. Driskill, 1981; Frankel, 1979; Haache and Townsend, 1981; Demery, 1984), by examining the properties of reduced-form exchange rate equations derived from structural models of the Dornbusch type. The empirical results the authors report are unfavourable to the model. However there are a number of problems with these tests, the main one being that they all assume the regressive form for expectations, which is in general incorrect (Minford and Peel, 1983). For example, Haache and Townsend (1981) and Frankel (1979, 1982) specify models in which lagged adjustment or wealth effects are introduced into the demand for money function or lagged adjustment is introduced into the interest arbitrage condition; in these specifications, expectations will not be regressive. Box 14.1 shows that overshooting also can, but does not necessarily, occur in equilibrium models.

Box 14.1

EXCHANGE RATE OVERSHOOTING IN EQUILIBRIUM MODELS

We now examine whether exchange rate overshooting, which occurs in the Dornbusch model as a result of disequilibrium in goods or labour markets, must always be regarded in a real economy as occurring as a consequence of such features. We will demonstrate that this is not the case. Overshooting can indeed occur in equilibrium open economy models. In order to explain this, we adopt the model

of the previous section.
In keeping with the equilibrium framework, we assume that all agents form expectations on the basis of the same set of macro information. This we date at $t-1$, and for simplicity ignore current global information (but this does not alter our point).
Hence the exchange market equilibrium condition becomes:

$$E_{t-1}S_{t+1} - S_t = R_t \tag{1}$$

and the real interest rate differential, r, is defined as:

$$r_t = R_t - E_{t-1}p_{t+1} + p_t \tag{2}$$

(We have for simplicity set the foreign real interest rate to zero.)
Demand for money is:

$$m_t = p_t + y_t - \lambda R_t \tag{3}$$

Money supply is:

$$\Delta m_t = \epsilon_t + \Delta \overline{m}_t + \Delta v_t \tag{4}$$

where

$$\Delta \overline{m}_t = u_t \tag{5}$$

This money supply function now allows not only for (unanticipated) temporary changes in the level of money (v) and once-for-all changes in the level (ϵ) but also for once-for-all changes in the steady state rate of increase (u). It will thus permit us to examine the different reactions to these shocks. (note $m_t = \sum_{i=0}^{\infty} \epsilon_{t-i} + \overline{m}_t + v_t$ where $\overline{m}_t = \sum_{i=0}^{\infty} u_{t-i}$)
From the definition of real interest rates and the real exchange rate, $x_t = S_t - p_t$, and from (1) we have:

$$(R_t - E_{t-1}p_{t+1} + p_t) = E_{t-1}S_{t+1} - E_{t-1}p_{t+1} - (S_t - p_t) = E_{t-1}x_{t+1} - x_t \tag{6}$$

i.e. the real interest differential must equal the expected real depreciation.
Now complete the model with an IS and Phillips curve:

$$y_t = -\alpha r_t + \delta x_t \tag{7}$$

$$y_t = \beta r_t + \gamma(p_t - E_{t-1}p_t) + \sigma y_{t-1} \tag{8}$$

Equation (8) has the full classical form discussed in chapter 3. Note that neither α nor β is infinite, so that r will vary. The model has been set up so that $y = r = x = 0$ in equilibrium. Notice that it belongs to the same Mundell-Fleming family as that of chapter 10. Define the superscript '^{ue}' as 'unanticipated at t-1'; hence for example $p_t^{ue} = p_t - E_{t-1}p_t$. Equations (6)–(8) can be solved as a recursive block in terms of p_t^{ue} to give:

$$x_t = \mu x_{t-1} + \pi_0 p_t^{ue} + (\pi_1 - \mu\pi_0)p_{t-1}^{ue} \tag{9}$$

where $\pi_0 = \frac{\gamma}{\alpha+\beta+\delta}, \pi_1 = \sigma\pi_0\frac{\alpha+\delta}{(\alpha+\beta)(1-\mu)+\delta}$ and μ is the (assumed unique) stable root of the characteristic equation

$$\mu^2 - (1 + \frac{\alpha\sigma+\delta}{\alpha+\beta})\mu + \frac{\sigma(\alpha+\delta)}{\alpha+\beta} = 0 \tag{10}$$

r and y have similar solutions: a first-order moving average in p^{ue} and first-order autoregressive coefficient μ. From (3) using these, we obtain:

$$p_t^{ue} = qm_t^{ue} \tag{11}$$

$$x_t^{ue} = \pi_0 qm_t^{ue} \tag{12}$$

where $q = \frac{1}{[1+\lambda+\lambda\pi_0+(\alpha+\delta)\pi_0]}$ is greater than 0 and less than 1. The nominal exchange rate depreciation is:

$$S_t^{ue} = x_t^{ue} + p_t^{ue} = q(1+\pi_0)m_t^{ue} \tag{13}$$

where $m_t^{ue} = \epsilon_t + u_t + v_t$
We can usefully rewrite

$$q(1+\pi_0) = \frac{\alpha+\beta+\delta+\gamma}{(\alpha+\beta+\delta)(1+\lambda)+\gamma(\lambda+\alpha+\delta)} \tag{14}$$

which makes it clear that the value is positive and greater or less than unity depending on all the impact parameters (however if $\lambda+\alpha+\delta$ is greater than 1 then it must be less). Notice that overshooting properties occur in the broad sense that both the nominal and the real exchange rate depreciate in a 'volatile' manner in response to positive money shocks.
For the case where the money supply is growing over time we define overshooting as a reaction of the nominal exchange rate by a greater

proportion than the change in the (current) equilibrium nominal exchange rate, i.e. that which would prevail were the present money supply difference to be maintained in perpetuity, apart from elements expected to be reversed (v_t).
On this basis, we can determine from (13) that:

1. The exchange rate may respond more than proportionately to a once-for-all change in m (ϵ_t), and so the equilibrium exchange rate. This is the overshooting considered by Dornbusch (l976), which deals with surprise shifts in the permanent level of m.
2. It also may respond more than proportionately to a rise in m which is due to a permanent rise in its growth rate (u), However, since there is no way that speculators can distinguish between ϵ and u shocks when they occur, the reactions to both are the same.
3. It also responds to a temporary change in m (v), which on our definition does not change the equilibrium exchange rate. This is also a form of overshooting, though not that dealt with by Dornbusch.

All these types of overshooting in response to monetary shocks are qualitatively the same as those in the 'sticky price' models of Dornbusch and Frankel, yet they emerge from an equilibrium model. By altering our assumptions about the availability of current global information, these results could be easily 'enriched' to give a variety of potential overshooting responses; substantial overshooting is exhibited in empirical application by an equilibrium model of the UK economy (Minford, l980, the Liverpool model). To sum up, volatility of the nominal exchange rate (overshooting), as well as of the real exchange rate and real interest differentials, is not prima facie evidence of 'price stickiness', 'disequilibrium' or 'inefficiency' in goods or labour markets.
Distinguishing Equilibrium from Disequilibrium Models of the Exchange Rate?
For good measure we can show that, on the basis of a reduced form exchange rate equation on its own, it is not possible to determine whether it comes from an equilibrium or disequilibrium model. For this purpose, we set up two models identical in all respects except in their 'supply' behaviour.
The first model is the one just dealt with; it consists of the demand for money function (3), money supply function (4), efficient market

condition (1), and IS curve (7) and its supply curve is an equilibrium one (8). The second model consists of the same equations apart from (8) where it has a sticky price Phillips curve like Frankel's, namely:

$$p_t - p_{t-1} = bx_{t-1} + \Delta\overline{m}_t \tag{15}$$

(We also assume in the spirit of disequilibrium models that speculators have full current information, and condition the expectations operator throughout on the basis of current information.) It turns out that the solution for x_t in this model is:

$$x_t = \mu_2 x_{t-1} + \frac{[\epsilon_t + \lambda u_t + (1-\mu_1^{-1})(v_t - v_{t-1})]}{\mu_2(\lambda+\alpha)} \tag{16}$$

where μ_2 is the stable, μ_1 the unstable root of the characteristic equation:

$$\mu^2 - \{2 - \frac{\lambda\gamma - \delta}{\lambda+\alpha}\}\mu + 1 - \frac{\gamma[1+\lambda]+\delta}{\lambda+\alpha} = 0 \tag{17}$$

Compare this to the solution for the equilibrium model:

$$x_t = x_{t-1} + \pi_0 q m_t^{ue} + q(\pi_1 - \pi_0) m_{t-1}^{ue} \tag{18}$$

This shows that it is not possible to distinguish between the equilibrium and disequilibrium models on the basis of the reduced-form (real) exchange rate equations alone; both are ARMA(1,0,1) time-series models (see Time Series Annex). It follows that the models can only be distinguished, if at all, on the basis of full structural estimation. This is another example of 'observational equivalence' (see chapter 15).

EMPIRICAL EVIDENCE ON MARKET EFFICIENCY

Our interest here is in the empirical evidence that testing for market efficiency sheds on the rational expectations hypothesis. In these tests, either part of the joint hypothesis may fail: the model of equilibrium expected returns or the RE hypothesis. But this is unavoidable in testing a hypothesis about expectations which are not directly observable.

Modelling of the equilibrium expected return is clearly crucial in empirical tests of market efficiency. As the examples above have illustrated asset prices or returns can have predictable patterns without necessarily violating the efficient markets hypothesis. The crucial element is that abnormal returns should not be systematically predictable.

Since a property of rational expectations is that any difference of outcome from the expected outcome is unforecastable from available information, we have that :

$$R_{it} = E_t R_{it} + v_{it} \tag{81}$$

so that v_{it} is independent of $E_t R_{it}$, the rational expectation. Substituting the determinants of $E_t R_{it}$from CAPM, for example, gives:

$$R_{it} = R_t + q_{it} + v_{it} \tag{82}$$

where q_{it} is the risk premium. If q_{it} were constant (82) can be estimated by ordinary least squares (with the coefficient on R_t constrained to unity), and the estimated error term, $\widehat{v_{it}}$, should be independent of all information available at the beginning of period t: not merely past $\widehat{v_{it}}$ (weak-form), but also all relevant data (such as money supply, inflation and growth). This is known as an orthogonality test. It is also possible to estimate (81) freely, in which case the coefficient on R_t should not be significantly different from unity, a further check on the joint hypothesis: the other tests apply as before. A further implication of (81) is that any trading rule, TR (a systematic rule for trading assets), which uses information at the start of t, including past $\widehat{v_{it}}$, to buy and sell asset i, intending to make profits because

$$E[R_{it} - E_t R_{it}] \mid TR > 0 \tag{83}$$

must fail under the efficiency hypothesis since by (81)

$$E[R_{it} - E_t R_{it}] = 0 \tag{84}$$

Since under the efficiency assumption any trading strategy has an expected abnormal return of zero it must do worse, given the increased transactions costs associated with an active rule compared with a trading strategy of buy and hold (do nothing). The returns to each rule differ by expected transactions costs.

An important class of models in which efficiency prevails (abnormal returns are unforecastable) is that of martingales. Formally if a variable Z_t is described by the stochastic process

$$E[Z_{t+1} \mid Z_{t,} Z_{t-1}, \ldots = Z_t \tag{85}$$

or

$$E[Z_{t+1} - Z_t \mid Z_{t,} Z_{t-1}, \ldots] = 0 \tag{86}$$

it is known as the martingale property.

The process

$$Z_{t+1} = Z_t + u_{t+1} \tag{87}$$

is therefore a martingale. The error term has the property that $E_t u_{t+1} = 0$. The error can exhibit structure such as time varying heteroskedasticity (e.g. ARCH effects – Time-Series Annex). If a constant term is added to (87) it is known as a submartingale. The terminology of random walk and martingale are often interchanged. In fact a random walk as conventionally defined is a stronger concept in that the error term is assumed to be independently and identically distributed (iid).

So the process

$$Z_{t+1} = Z_t + \alpha + u_{t+1} \tag{88}$$

where α is a constant and $E_t u_{t+1} = 0$, is a submartingale and a random walk if u_{t+1} is iid.

THE RATIONALE FOR MARKET INEFFICIENCY

Broadly the implicit mechanism involved for the market to be efficient in the sense of Fama is that if asset markets were not efficiently aggregating and processing information the disparity between fundamental values and market prices would present traders with profit opportunities. Rational speculators it is argued will, essentially instantaneously, drive asset prices back to their fundamental values.

The case for market inefficiency rests on either or both of the premisses that prices of assets move when there is no new information concerning fundamentals and that the action of speculators may not move asset prices towards their fundamental values.

Some empirical observations that are offered as consistent with these premisses are market crashes such as October 19th 1987 when world markets fell around 20 per cent without apparently any important news being evident which could account for the fall, or sustained rises in prices, known as bubbles, that are seemingly inexplicable in terms of market fundamentals.

Also it is observed that many trading decisions are based on past prices. Chartism, the extrapolation of past prices, is widely employed as the basis for trading rules. In addition, there is extensive use of stop-loss orders, whereby an asset is sold if its price falls by a pre-specified amount as well as the growth of dynamic hedging strategies such as portfolio insurance where investors buy (sell) into rising (falling) markets. Whilst none of these activities necessarily imply market inefficiency they raise the question as to whether informed rational speculators are arbitraging out the trading rules of any uninformed or irrational traders; if they are not, then inefficiency may be present.

The first theoretical model of inefficiency we examine is that of rational bubbles.

SPECULATIVE BUBBLES

History is replete with examples where asset prices have exhibited dramatic increases then falls which it is argued are not readily explained by movements in the 'fundamentals' of the asset such as the expected future dividend stream. One example is the South Sea bubble in the UK: in the 18th century the stock of a company which traded in the South Sea experienced exponential price increases before plummeting in value. Another is the tulip bulb episode in Holland in the late sixteenth century where (single) tulip bulbs were exchanged for land and gold before tulips ultimately became near worthless (though see Garber, 1989).

It is argued that these episodes represent speculative bubbles in which the anticipation of future capital gains leads to spiralling upward price movements before the bubble eventually collapses or pops and the price exhibits dramatic falls. Since the 1980s there has been a considerable amount of theoretical and empirical work on speculative bubbles. Our purpose in this section is to provide an introduction to this literature; our discussion is related to our earlier one in chapter 2 where we showed that rational expectations with expectations of future variables can have bubbles in their solution.

We assume for simplicity that investors are risk neutral so that via arbitrage expected returns are equal to those on a riskless asset with rate of return $\overline{R}$, assumed constant, so that $E_t R_{t+1} = \overline{R}$. Given these assumptions we obtain (31) above, which we reproduce for convenience:

$$P_t = \frac{1}{(1+\overline{R})} E_t P_{t+1} + \frac{1}{(1+\overline{R})} E_t D_{t+1} \tag{89}$$

Assuming rational expectations and solving this model forwards N periods as in 33 we obtain

$$P_t = E_t \left[\sum_{i=1}^{N} \frac{D_{t+i}}{\left(1+\bar{R}\right)^i} \right] + \frac{E_t P_{t+N}}{(1+\overline{R})^i} \tag{90}$$

The second term in (90) is the discounted value of the stock price N periods in the future. In the absence of bubbles as we let N go to infinity we assume that this term goes to zero. If so, the current asset price is equal to the expected value of the stream of dividends into the indefinite future: this expression is the fundamental of the process, F_t.

The idea of speculative bubbles is that equation (90) is also consistent with rational expectations solutions other than the fundamental solution. If we try the solution

$$P_t = F_t + B_t \tag{91}$$

in (89) we find that as long as B_t follows the process

$$(1+\overline{R})B_t = E_t B_{t+1} \tag{92}$$

it is a valid mathematical solution to (89).

By substitution of (91) in (89) we obtain

$$P_t = F_t + B_t = \frac{1}{(1+\overline{R})}E_t B_{t+1} + \frac{1}{(1+\overline{R})}E_t F_{t+1} + \frac{1}{(1+\overline{R})}E_t D_{t+1}$$

(92) is consistent with this and the solution form for

$$F_t = \frac{1}{(1+\overline{R})}E_t F_{t+1} + \frac{1}{(1+\overline{R})}E_t D_{t+1}$$

is the same as (90) where $\frac{E_t F_{t+N}}{(1+\overline{R})^i}$ tends to zero as N tends to ∞. In other words if agents believe that the process B_t is driving asset prices then it will be a "rational" solution, a "self-fulfilling prophecy"; notice that it implies $E_t P_{t+N} = E_t F_{t+N} + E_t B_{t+N}$ which explodes endlessly, so that the bubble cannot be expected to burst ever.

As a consequence of the rational expectations assumption

$$B_{t+1} = E_t B_{t+1} + \epsilon_{t+1} \tag{93}$$

where ϵ_{t+1} is a random error.

Substitution of (92) into (93) gives

$$B_{t+1} = (1+\overline{R})B_t + \epsilon_{t+1} \tag{94}$$

(94) makes clear that the deterministic solution for B is an asymptotically explosive process. It is therefore important that there be no transversaility condition putting a limit on this process; such a condition was how we ruled out bubbles in chapter 2.

We can also write the solution for the bubble in the form (see Salge, 1997)

$$B_t = \frac{M_t}{\alpha^t} \tag{95}$$

where $\alpha = \frac{1}{(1+\overline{R})}$ and

$$E_t M_{t+1} = M_t \tag{96}$$

implying that M_t is a martingale process.

Leading (95) one period and taking expectations, noting that $\frac{1}{\alpha^t}$ is a deterministic process, we obtain

$$E_t B_{t+1} = \frac{E_t M_{t+1}}{\alpha^{t+1}} = \text{from (96)}\ \frac{M_t}{\alpha^{t+1}} = \frac{\alpha^t B_t}{\alpha^{t+1}} = \frac{B_t}{\alpha} \tag{97}$$

Consequently a bubble contains a martingale component and any process for M_t satisfying this condition is a valid rational bubble.

A general form of a stochastic martingale process is given by

$$M_t = \rho_t M_{t-1} + u_t v_t \tag{98}$$

where the random variable ρ_t, has conditional expectation that $E[\rho_{t+1} \mid I_t] = 1$. Other assumptions are that $E\{v_{t+i}\} = 0,\ i = 1-n,\ ,\ E[\rho_t M_t \mid I_t] = 0,\ E[\rho_t v_t \mid I_t] = 0,\ E[\rho_t u_t \mid I_t] = 0,\ E[u_t M_t \mid I_t] = 0,\ E[u_t v_t \mid I_t] = 0,\ E[v_t M_t \mid I_t] = 0$.

These assumptions establish the independence of all stochastic variables in all leads and lags. No restricting assumptions are required concerning the nature of the random variable u_t.

Suppose $u_t = 1$; define a random variable h_t that is normally distributed with mean μ and variance σ_h^2. We can then exploit the property of conditional log normality and define ρ_t for any arbitrary constant λ as

$$\rho_t = e^{\left[\lambda h_t - \left(\lambda\mu + \frac{\lambda^2 \sigma_h^2}{2}\right)\right]} \tag{99}$$

because

$$E e^{\lambda h_t} = E e^{\lambda h_{t+1}} = e^{\lambda\mu + \frac{\lambda^2 \sigma_h^2}{2}} \tag{100}$$

Taking expectations of (99) the condition $E\rho_{t+1} = 1$ is satisfied.

Consequently we can write the martingale process of (98) as:

$$M_t = e^{\left[\lambda h_t - \left(\lambda\mu + \frac{\lambda^2 \sigma_h^2}{2}\right)\right]} M_{t-1} + v_t \tag{101}$$

Substitution of (101) into (95) gives the implied bubble as:

$$B_t = e^{\left[\lambda h_t - \left(\lambda\mu + \frac{\lambda^2 \sigma_h^2}{2} + \ln\alpha\right)\right]} B_{t-1} + \epsilon_t \tag{102}$$

where $\epsilon_t = \frac{v_t}{\alpha^t}$ and $e^{-\log\alpha} = \frac{1}{\alpha}$.

($E_t\epsilon_{t+i} = 0$ for $i \geq 1$. Note substitution of M_t from (95) into (98) (with $u_t = 1$) gives $B_t = \alpha^{-1}\rho_t B_{t-1} + \frac{v_t}{\alpha^t}$; employing the defintion of ρ_t from (99) we have (102)).

Several different types of bubbles can be obtained exploiting the insight that rational bubbles include a martingale process. In fact any process that follows a martingale process can be included in the bubble.

For example suppose the fundamental can be described by a martingale process

$$F_t = F_{t-1} + \theta_t \tag{103}$$

where θ_t is a random variable.

In this case we can let M_t be:

$$M_t = F_t \text{ so that the bubble is } B_t = \frac{F_t}{\alpha^t} \tag{104}$$

For more general specifications of fundamentals the method is to substitute appropriate random variables for h_t. For example let us add a constant component to fundamentals so that they follow the process

$$F_t = F_{t-1} + \mu + \eta_t \tag{105}$$

where η_t is $N(0,\ \sigma_h^2)$. In order to find the bubble that corresponds to this process we inspect the general bubble formulation (102). We can substitute for the variable h_t in this formulation so long as the process we substitute in has a mean of μ and a variance σ_h^2, because for these parameter values we can obtain the martingale process (101). $F_t - F_{t-1}$ in (105) has a constant mean (μ) and variance σ_h^2 so we can substitute $F_t - F_{t-1}$ for h_t into the general bubble formulation (102) (with $\epsilon_t = 0$ assumed zero here for simplicity) to obtain

$$\frac{B_t}{B_{t-1}} = e^{\left[\lambda(F_t - F_{t-1}) - \left(\lambda\mu + \frac{\lambda^2\sigma_h^2}{2} + \ln\alpha\right)\right]} = \frac{e^{\lambda F_t - \left(\lambda\mu + \frac{\lambda^2\sigma_h^2}{2} + \ln\alpha\right)t}}{e^{\lambda F_{t-1} - \left(\lambda\mu + \frac{\lambda^2\sigma_h^2}{2} + \ln\alpha\right)(t-1)}} \tag{106}$$

(106) implies that the bubble process in fundamentals corresponding to the fundamental process (105) is given by

$$B_t = e^{\left[\lambda F_t - \left(\lambda\mu + \frac{\lambda^2\sigma_h^2}{2} + \ln\alpha\right)t\right]} = e^{\left[\lambda F_t - \left(\lambda\mu + \frac{\lambda^2\sigma_h^2}{2}\right)t - \ln\alpha^t\right]} \tag{107}$$

If fundamentals follow the geometric process

$$\ln F_t - \ln F_{t-1} = \mu + \eta_t \tag{108}$$

where $\eta_t = N(0,\ \sigma_h^2)$, we obtain following the above procedure

$$\frac{B_t}{B_{t-1}} = e^{\left[\lambda(\ln F_t - \ln F_{t-1}) - \left(\lambda\mu + \frac{\lambda^2\sigma_h^2}{2} + \ln\alpha\right)\right]} = \frac{e^{\lambda \ln F_t - \left(\lambda\mu + \frac{\lambda^2\sigma_h^2}{2} + \ln\alpha\right)t}}{e^{\lambda \ln F_{t-1} - \left(\lambda\mu + \frac{\lambda^2\sigma_h^2}{2} + \ln\alpha\right)(t-1)}} \tag{109}$$

which implies

$$B_t = e^{\lambda \ln F_t - \left(\lambda\mu + \frac{\lambda^2\sigma_h^2}{2} + \ln\alpha\right)t} = F_t^{\lambda} e^{-\left[\left(\lambda\mu + \frac{\lambda^2\sigma_h^2}{2}\right)t - \ln\alpha^t\right]} \tag{110}$$

(note $e^{\lambda \ln F_t} = F_t^{\lambda}$). We observe that if $\lambda\mu + \frac{\lambda^2\sigma_h^2}{2} + \ln\alpha = 0$, then the bubble process is given by

$$B_t = F_t^{\lambda} \tag{111}$$

Alternatively if $\lambda = 0$

$$B_t = \frac{1}{\alpha^t} \tag{112}$$

When a bubble depends on its own value in the previous period it is called a Markovian bubble. When the bubble depends on fundamentals it is called an intrinsic bubble (Froot and Obstfeld, 1991). When the bubble depends on arbitrary processes it is called an extrinsic or extraneous bubble.The solution for extraneous bubbles follows the same procedure as for intrinsic bubbles. For instance if the extraneous process, S_t, follows a martingale process then the rational bubble is given by $B_t = \frac{S_t}{\alpha^t}$. Salge (1997) demonstrates how to solve for bubbles for the general ARMA specification of fundamentals or extraneous variables. Depending upon the specification of λ a variety of different bubble processes are feasible.

If bubbles were non-stochastic, $\epsilon_{t+1} = 0$, then the solution to (92) is simply

$$B_t = B_0(1 + \overline{R})^t \tag{113}$$

so that the solution for the price would simply embody a deterministic explosive component and would in principle be readily amenable to statistical tests.

Whilst the possibility of such deterministic bubbles have been investigated empirically for periods of hyperinflation before monetary reform (Flood and Garber, 1980), it would appear that bubbles of a deterministic form are not features of asset or other prices. As a consequence bubbles would not apear to be plausible empirically unless there is a significant probability that they will collapse after reaching high levels.

Blanchard (1979) and Blanchard and Watson (1982) proposed a probabilistic bubble that can embody this feature. Essentially there are two regimes which occur with constant probabilities q and $1 - q$. In the first state (A) the bubble survives with probability q and continues to increase at an expected rate of $E_t B_{t+1}/A = \frac{(1+\overline{R})B_t}{q}$. In the second state ($C$), with probability $1 - q$, the bubble collapses so that $E_t B_{t+1}/C = 0$.

Adding stochastic terms we have

$$B_{t+1} = \frac{(1+\overline{R})B_t}{q} + \epsilon_{t+1} \text{ with probability } q \text{ in state } A \tag{114}$$

and:

$$B_{t+1} = \epsilon_{t+1} \text{ with probability } (1-q) \text{ in state } C \tag{115}$$

where the error has the property that $E_t \epsilon_{t+1} = 0$ As a consequence we have

$$E_t B_{t+1} = q\left(\frac{(1+\overline{R})B_t}{q}\right) + (1-q)0 \tag{116}$$

We note from (116) that the bubble has the form of equation (92) and is therefore a valid rational expectations solution. In addition we observe that in state A the bubble grows at a faster rate on average $(\frac{1}{q})$ than the non-probabilistic bubble in order to compensate for the probability of collapse. Collapsing bubbles are consistent with the analysis employing martingales; we simply define the process

$$M_t = \frac{\rho_t}{\pi} M_{t-1} + u_t v_t \text{ with probability } \pi \tag{117}$$

and

$$M_t = u_t v_t \text{ with probability } 1 - \pi \tag{118}$$

We can also derive the expected excess return, which is defined as the return on the asset which incorporates the bubble minus the rate of return on the riskless asset in regime A and C. Defining the expected excess return at the end of period one as $X_{t+1} = E_t R_{t+1} - \overline{R}$, then from (89)

$$X_{t+1} = \frac{E_t P_{t+1} + E_t D_{t+1}}{P_t} - (1+\overline{R}) \tag{119}$$

Taking expectations at time t of (91) at time $t+1$ we obtain

$$E_t P_{t+1} = E_t F_{t+1} + E_t B_{t+1} \tag{120}$$

where

$$E_tF_{t+1} = E_t\left[E_{t+1}\left\{\sum_{i=1}^{\infty}\frac{D_{t+i+1}}{(1+\overline{R})^i}\right\}\right] = E_t\left\{\sum_{i=1}^{\infty}\frac{D_{t+i+1}}{(1+\overline{R})^i}\right\} \quad (121)$$

now

$$E_tF_{t+1} + E_tD_{t+1} = E_t\sum_{i=1}^{\infty}\frac{D_{t+i+1}}{(1+\overline{R})^i} + E_tD_{t+1} = (1+\overline{R})F_t = (1+\overline{R})(P_t - B_t) \quad (122)$$

so substituting (122) and (120) into (119) we obtain

$$X_{t+1} = \frac{(1+\overline{R})(P_t - B_t) + E_tB_{t+1}}{P_t} - (1+\overline{R}) \quad (123)$$

In regime A we substitue for E_tB_{t+1} from (116) into (123) to obtain expected excess returns as

$$X_{t+1} = \frac{(1+\overline{R})B_t(1-q)}{qP_t} \quad (124)$$

In regime C we substitute for E_tB_{t+1} from (116) into (123) to obtain expected returns as

$$X_{t+1} = -\frac{(1+\overline{R})B_t}{P_t} \quad (125)$$

We observe from (124) and (125) that expected excess returns differ substantially in the two periods. However expected excess returns are zero across the two regimes. Rational bubbles do not create a predictable pattern in excess returns; rather they create volatility in prices.

Whilst the the above probabilistic bubble has the property of collapsing, Diba and Grossman (1988) show that the impossibility of negative rational bubbles in stock markets (because of the implication of a negative expected asset price in the future which violates limited liability) implies that a bubble once collapsed can never restart. Essentially if the bubble takes a zero value, then from (92) its expected value in the future is zero. Because negative values are ruled out for stock prices this implies that the average of the positive values must be zero, but this is a contradiction, so it can only be zero if it takes the value of zero in all future periods. This point also carries the implication that if a bubble exists today it must have always existed, i.e. since the moment trading began.

Evans (1991) and Van Norden (1996) have formulated processes for bubbles that meet this theoretical point. Van Norden allows for the possibility that the bubble is expected to collapse only partially in state C and that the probability of a bubble's continued growth falls as the bubble grows

Van Norden specifies that the probability of the bubble's continued growth is given by

$$q_t = q(B_t) \text{ with } \frac{dq(B_t)}{d \mid B_t \mid} < 0 \tag{126}$$

(the absolute value of the bubble to allow for negative bubbles in markets such as exchange rates.).

In regime C the bubble is expected to collapse only partially and is given by

$$E_t B_{t+1} = u(B_t) \text{ in state } C \text{ with probability } q_t \tag{127}$$

where $u(.)$ is a continuous and everywhere differentiable function such that $u(0) = 0$ and $1 \geq \frac{du(B_t)}{dB_t} \geq 0$.

In state A the bubble is expected to grow at rate

$$E_t B_{t+1} = \frac{(1+\overline{R})B_t}{q(B_t)} - \frac{[1-q(B_t)]u(B_t)}{q(B_t)} \text{ with probability } 1-q_t \tag{128}$$

It is easy to verify that (128) and (127) imply equation (92). We can deduce from (128) that the expected value of the bubble in the surviving state is a decreasing function of the probability of survival $q(B_t)$. As a consequence, as with the Blanchard bubble, the greater the probability of collapse the larger is the expected gain on a surviving bubble in order to compensate the investor for the possibility of collapse.

Van Norden's specification of the bubble process is interesting in demonstrating how the probability of collapse can be readily endogenised, and also how the process can be modified so that the bubble is not necessarily zero in the collapsed state. Both these features could be relevant in empirical work on detecting bubbles.

Evans formulates a bubble that is always positive but nevertheless periodically collapses. The Evans bubble takes the form

$$B_{t+1} = (1+\overline{R})B_t u_{t+1} \text{ if } B_t \leq k \tag{129}$$

and

$$B_{t+1} = \left[\delta + \frac{\theta_{t+1}(1+\overline{R})[B_t - (1+\overline{R})^{-1}\delta]}{q}\right] u_{t+1} \text{ if } B_t > k \tag{130}$$

where k and δ are positive parameters with $0 < \delta < (1+\overline{R})k$ and u_{t+1} is an exogenous independently and identically distributed (iid) positive random variable with $E_t u_{t+1} = 1$. θ_{t+1} is an exogenous i.id Bernoulli process(independent of u) which takes the value unity with probability q and zero with probability $1-q$.

Taking expectations at t of (129) we observe that the bubble is equation (92) when it is in the regime where $B_t \leq k$. When $B_t > k$ the bubble has the expected value of

$$E_t B_{t+1} = \left[\delta + \frac{(1+\overline{R})[B_t - (1+\overline{R})^{-1}\delta]}{q}\right] = \left[\frac{(1+\overline{R})B_t}{q} - \frac{\delta(1-q)}{q}\right] \text{ with probability } q \quad (131)$$

and

$$E_t B_{t+1} = \delta \text{ with probability } 1-q \quad (132)$$

so that when $B_t > k$

$$E_t B_{t+1} = q\left[\frac{(1+\overline{R})B_t}{q} - \frac{\delta(1-q)}{q}\right] + (1-q)\delta = (1+\overline{R})B_t \quad (133)$$

Consequently the bubble satisfies condition (92).

The Evans bubble has the property that when $B_t \leq k$ it grows at mean rate $1+\overline{R}$. When $B_t > k$ the bubble 'erupts' and grows at a faster mean rate as long as the process continues. When the bubble collapses it falls to a mean value of δ and the bubble process begins again. Evans notes that by varying the parameters δ, k and q one can create bubbles where the frequency with which the bubble erupts, the average length of time before collapse and the scale of the bubble vary. An example of an Evans bubble is depicted in Figure 14.3.

An important feature of the Evans bubble is that standard empirical tests for bubbles based on unit root and cointegration methods (Time-Series Annex) may not detect bubbles when they are present. Evans simulated bubbles and applied unit root tests to the simulated bubbles. Even though bubbles are asymptotically explosive for $q \leq 0.75$ for more than 90% of the simulated bubbles of length 100 observations, statistical tests for unit roots rejected the hypothesis of a bubble in favour of a stable alternative.

More recent empirical work has endeavoured to apply alternative econometric approachs such as switching regime regression models to ascertain the presence of bubbles (Hamilton, 1994; Van Norden 1996; Van Norden and Vigfusson, 1998) given that the Evans-type model describes behaviour for different regimes. These statistical methods appear to offer greater promise of detecting bubbles. As yet however it is probably fair to say that the different empirical methods which have been applied to a

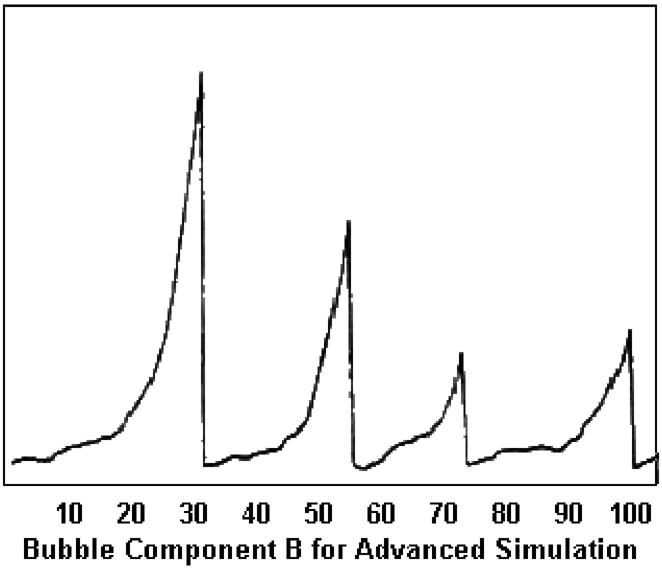

Figure 14.3: Example of an Evans Bubble

variety of different asset markets including stock and exchange markets have not generated any consistent evidence of speculative bubbles.

It should be noted that theoretical work by Tirole (1982, 1985) has demonstrated that bubbles cannot arise in certain models. For example in an overlapping generations model with an infinite number of finite-lived representative agents, Tirole shows that when the interest rate exceeds the growth rate in the economy (so that it is dynamically efficient) a bubble would ultimately violate some agents' budget constraints. It would appear that bubbles could only exist in economies which are dynamically inefficient if the model is of the representative-agent type and agents are rational. From this perspective Froot and Obstfeld (1991) suggest interpreting empirical tests for bubbles as tests of the rational expectations assumption.

We can put this another way. RE models are, generally, 'over-identified'. That is, everything in the reduced form is derived from structural relationships and more restrictions exist than reduced-form parameters; the reduced-form errors are combinations of the structural parameters and structural errors. Errors such as bubbles are not present in the structural model and hence should not appear in the reduced form.

Only fundamentals are included in it.

Should we write down a structural model in which some bubble variable is allowed to enter, then we must ensure that it satisfies both the transversality conditions on the model (including private and budget constraints as they tend to infinity) and the optimality conditions. For example, as we showed in chapter 2, deterministic explosive bubbles in prices will violate reasonable government limits on inflation and so will be ruled out from the start. In the Blanchard and Watson example reviewed above, of a stochastic bubble, at least one of the paths must have a positive probability of continuing indefinitely for the bubble even to begin. Is such a path consistent with such limits? Above we noted the objections of Tirole to bubbles in one context, and the caveats expressed by Obstfeld and Rogoff. In each case of a bubble we must ask whether theory as above permits it.

When we turn to empirical testing, we can note that the 'forward solution' of the model discussed in chapter 2 has the form of a bubble. One can for example write down a model of prices in which a large positive future shock (to the money supply, say) is expected with some probability; as the period approaches the price level rises until at the period it either jumps upwards if the shock occurs or collapses if it does not. Distinguishing this non-bubble solution from a supposed bubble clearly is difficult. Remember a bubble relates to what people in markets expect for the future; this could well relate to an expectation about a fundamental.

The difficulty with bubbles is therefore both theoretical and empirical. This critique of bubbles within rational expectations models does not of course extend to models where expectations are not rational; here, by construction, people may believe in curiosa – 'fads' and so forth. To such models we now turn.

Fads and Noise Traders

Shiller (1984), (1989) and (1997) has emphasised the importance of mass psychology in financial markets with the implication that investors may exhibit fashions or fads. A fad is a depature from fundamental values due to 'psychologically' induced changes in market sentiment. We first illustrate the implications following the model in Fama and French (1988) and Cutler, Poterba and Summers (1990, 1991).

Let

$$P_t = F_t + e_t \qquad (134)$$

where P_t is the return or price of an asset, F_t is the fundamental value and e_t is a fad. Fundamentals are assumed to follow the process

$$F_t = F_{t-1} + h_t \tag{135}$$

The fad is described by the stationary process

$$e_t = \rho e_{t-1} + v_t \tag{136}$$

where ρ is a positive constant less than unity and v_t is a random error.

Cutler, Poterba and Summers (1991) consider the case where a proxy, F_t^*, is available to measure fundamentals, given that in applied work we typically do not measure the fundamentals with precision or are employing approximations to the theoretical model. Let

$$F_t = F_t^* + w_t \tag{137}$$

where w_t is the measurement error assumed to be random.

Differencing (134) and substituting from (135) and (136) we obtain the true reduced form for ΔP_t as

$$\Delta P_t = (\rho - 1)(P_{t-1} - F_{t-1}) + h_t + v_t \tag{138}$$

Consider the regression estimate of β in the relationship

$$\Delta P_t = \alpha + \beta(P_{t-1} - F_{t-1}^*) + \theta_t \tag{139}$$

The regression coefficient will have a probability limit of

$$\beta = \frac{(\rho - 1)Cov(P_{t-1} - F_{t-1}, P_{t-1} - F_{t-1}^*)}{var(P_{t-1} - F_{t-1}^*)} = \frac{(\rho - 1)Cov(e_{t-1}, e_{t-1} + w_{t-1})}{Var(e_{t-1} + w_{t-1})} = \frac{(\rho - 1)var(e)}{var(e) + var(w)} \tag{140}$$

The implication is that changes in asset prices will be predicted by lagged deviations of the price from its proxy fundamental (for which in their empirical work Cutler, Poterba and Summers (1991) employ the log of the real dividend). In addition prices or returns can be shown to exhibit a predictable pattern.

Of course predictable patterns in the difference between prices and fundamentals raises the issue of why rational speculators do not arbitrage the process. It is implicitly assumed that they face some sort of liquidity constraint preventing them from dominating the market. However, it is also possible that they do so arbitrage, so that the fad occurs but is then removed from the data next period. This would correspond to $\rho = 0$

in our example. In this case, prices could differ from their fundamental value so implying inefficiency, and yet the difference between the price and fundamentals could still be unpredictable. This distinction points to a caveat in the interpretation of efficiency given by Fama; Shiller makes the point that the absence of predictability does not necessarily imply market efficiency. Clearly the existence of systematic abnormal profits is sufficient to invalidate efficiency. However absence of predictable abnormal returns could be consistent with market inefficiency if prices were more volatile than was consistent with the volatility in the fundamental. Bubbles or some fad processes are consistent with this interpretation of inefficiency. Efficiency therefore requires the complete elimination of fads from the price process ($e_t = 0$); otherwise a trading rule operating on $P_t = F_t$ would make money (this is a particularly simple rule, viz. of current arbitrage).

Other models of market inefficiency assume the coexistence of heterogeneous traders. In these models there are smart traders, the rational speculators, as well as the noise traders. The noise trader trades on fads or charts or other extraneous information ('noise'). The key requirement of the models is that arbitrage activity by the smart traders is limited. This can be a consequence of the assumption of risk-averse rational speculators combined with a micro-structure rationale where for example arbitrageurs have limited capacity to borrow funds due to signalling problems (see e.g. Shleifer and Vishny, 1997).

The consequence is that arbitrage activity by the rational traders may not eliminate the influence of the noise traders from the outcome for returns or prices so that the market is inefficient in aggregate (see e.g. Figlewski, 1979).

Other specifications of this type of model have the informed traders as not holding rational expectations but rather forming their expectations on the basis of a relatively well-informed rule such as PPP deviations for exchange rates (section below) or price-dividend deviations for stock prices using say the Gordon (1962) model . If it is assumed that the proportion of the two types of agent in the market varies with the extent of the deviation of the asset price from its long run value, then these models can exhibit complex nonlinear dynamics of the asset price including chaotic outcomes (Time-Series Annex; and e.g. De Grauwe, Dewachter and Embrechts, 1993).

Some recent models of asset price determination investigate more formally the micro-structure determining asset prices. An important class of models consider the implications of assuming that traders do not behave competitively and take prices as given. The models are solved using the game-theoretic Bayesian Nash equilibrium concept where the strategies of other traders rather than prices are taken as given. These models allow analysis of strategic interactions in which traders take their price impact into account (see Brunnermeier, 1999, for a survey). In essence each trader recognises that large trades will move prices against him. The models incorporate features such as asymmetric information, noise traders, speculators, market-makers and explicit assumptions concerning the mechanism by which traders submit orders to trade. They produce many interesting insights.

When it is assumed that traders have short investment horizons, perhaps due to financial constraints which make arbitrage activities cheaper for short-term assets, herd behaviour can result (see e.g. Froot, Scharfstein and Stein, 1992). In addition strategic interaction can give rise to rational profit-maximising speculators, accentuating asset price volatility caused by noise traders as they strategically exploit feedback elements such as herding (see e.g. De Long, Shleifer, Summers and Waldman, 1991).

The models can also generate market crashes as in e.g. Romer (1993) where information is heterogeneous and investors are uncertain about the quality of information other investors have.

Although this literature is rich in insight and can provide a rationale for market inefficiency, it remains an empirical issue whether efficiency as defined by Fama is violated. We now turn to some empirical tests.

EMPIRICAL TESTS OF MARKET EFFICIENCY

There has been an enormous amount of empirical work which has examined the efficiency of asset markets including gambling markets (see e.g. Sauer, 1998; Thaler and Ziemba, 1987). Violations of the efficient markets hypothesis are typically referred to as anomalies. The anomalies reported may constitute evidence of abnormal returns or violations of the rational expectations assumption. Before considering some of this evidence it is important to note that the dangers of data-driven inference or data-snooping whilst always a consideration in applied analysis (see Leamer, 1978, for a discussion of pretest bias) needs to be given formal weight when interpreting some anomalies as demonstrated by Sullivan, Timmermann and White (1999, 2001). In the limited sample sizes typically encountered in studies, significant relationships or sytematic patterns are bound to occur if the data are analyzed with sufficient intensity. For example the orthogonality property of rational expectations will be rejected by the usual statistical criteria five percent of the time. Clearly if authors are searching in the universe of variables and significant rejections are reported it could appear that the hypothesis is invalid even though it is true.

An apparent rejection of the market efficiency hypothesis which is data-driven is the finding reported by numerous researchers that stock returns exhibit seasonal regularities (see Sullivan, Timmermann and White, 2001, for key references). Abnormal returns are related to day of the week, week of the month, month of the year, holidays etc. If these results have validity then the efficient markets hypothesis is in serious trouble given it is 'such a simple violation'.

The issue is one of 'data-snooping' — in other words, researchers cannot avoid being influenced by their sample of data. They 'find' rules that are in that set of data by repeated trials. This will occur even when some data is kept aside – because this data too is known about to some degree and trials on it may lead to respecification of the rule on the used data before a renewed trial. To control for this problem the researcher must not allow the data to influence his choice of rule. Thus all possible rules must be given an equal chance to work on the data; some of course will work well for particular samples even if they do not in general work at all. By doing this exercise for all rules for repeated samples one may calculate the probability of a rule 'doing well' in a given sample. Clearly for a rule to be considered significant it must do better than this. This exercise has been carried out by Sullivan, Timmermann and White (2001) for calendar effects.

They construct a universe of calendar trading rules, using permutational arguments that do not bias in favour or against particular calendar rules.They consider some 9500 calendar effects. They imagine that this set of calendar rules was not inspected by any one individual investor or researcher. Rather the search for calendar rules has operated sequentially across the investment community or researchers with the results of investors or researchers being reported through the survival of the 'fittest' calendar rules. Their findings are striking and important. They find that although many calendar rules generate abnormal returns that are highly significant when considered in isolation, once the dependencies operating across different calendar rules are allowed for, then the best calendar rule is no longer significant. In addition they consider a smaller number of 244 rules that remove any doubts that irrelevant rules are pooled with'genuine' rules in the 9500 experiment. Their results are the same. The apparent statistical significance of the best calendar effects is not robust to data-snooping effects.

Weak-form efficiency tests

Weak-form efficiency in the sense of Fama implies that the expected returns from a trading rule based on past prices should be less than those generated by a buy-and-hold strategy equal to the number of trades times the transactions costs. In fact according to Taylor and Allen (1992) and Brock, Lakonishok and LeBaron (1992) chartist or technical techniques are widely employed in asset markets as either a basis for published technical commentary or for direct use. The term technical analysis or chartism is given as a generic title for any trading rule based on the past history of prices or returns. Essentially more or less sophisticated rules are employed to extrapolate past changes in asset prices or returns. One common example is to employ two moving averages of past returns one based on a short horizon, say 1 to 10 days, and the other on a long moving average, say 150 to 500 days. When the short moving average rises above (or falls below) the long moving average this is a signal to buy (or sell). A variant of this procedure is to modify the rule by introduction of a band around the moving average to eliminate 'whiplash' signals when the two moving averages are close.

There is a plethora of empirical work that has investigated the efficacy of technical analysis — e.g. Alexander (1961); Fama and Blume (1966); Taylor and Allen (1992); Brock, Lakonishok and LeBaron (1992); Neely, Weller and Dittmar (1997); Sullivan, Timmermann and White (1999).

Whilst many rules developed by chartists have been reported to generate abnormal returns, and some theoretical analysis — e.g. Brown and Jennings (1989); Blume, Easley and O'Hara (1994) — demonstrates that technical analysis can have value particularly for small less widely-followed assets, the empirical results of Sullivan, Timmermann and White (1999) appear to be of importance in the interpretation of chartism. Following a similar methodolgy to the one outlined above they construct 7846 parameterizations of trading rules which are applied to the Dow Jones Industrial Average over the full period from 1897 to1986 as well as four subperiods. The period 1987-1996 is employed to evaluate the rules on a hold-out sample. Their idea is to develop a test statistic which evaluates the performance of a chartist rule relative to the full set of models that gave rise to the rule, so that the effect of data-snooping is explicitly allowed for. One important conclusion is that even though a particular trading rule is capable of producing superior performance of almost 10% during the sample period, which is significant at 4% level when considered in isolation, the fact that the trading rule is drawn from a wide universe of rules means that its effective data-snooping-adjusted probability value is 0.9, i.e. totally insignificant. They also find that the best-performing trading rule in sample is totally insignificant out of sample even at conventional statistical levels. Clearly chartist rules may exist which generate abnormal returns; however the analysis of Sullivan, Timmermann and White (1999) demonstrates that great care has to be taken in their interpretation.

Another issue concerning chartists or technical rules is that they can be interpreted as special cases of univariate time series of a linear or nonlinear form (Time-Series Annex). If there is predictability in past returns it would appear that employment of an explicit univariate model would in general dominate. Studies such as Diebold and Nason (1990), Lane, Peel and Raeburn (1996) are suggestive that nonlinear univariate models for modelling the conditional mean of a series provide improvement over linear models but do not appear to forecast better than linear models, with neither of them generating abnormal returns (though see Granger and Pesaran, 1999).The results of Hsieh (1991) who found that stock returns exhibit nonlinear structure but that this is parsimoniously captured by a model linear in mean but nonlinear in variance (Time-Series Annex) is probably the model that is currently thought of as most applicable to asset prices or returns. Clearly chartist methods may conceivably deal more parsimoniously with nonstationarities or regime changes than explicit nonlinear models though it is not altogether clear why this should be the case.

Overall it would appear that the market efficiency hypothesis has not as yet been invalidated at the weak-form level.

Semi-Strong Market Efficiency

Numerous empirical tests of semi-strong efficiency exist. Some of the work reporting anomalies would appear to be subject to the data-snooping caveat. Others are harder to explain. In the latter context we mention the empirical work that demonstrates that monthly or quarterly stock returns are predictable (e.g. Campbell, 1987; Fama and French, 1989; Pesaran and Timmermann, 1994, 1995, 2000). Consideration of the loglinear form of the present value model illustrates that if expectations of future dividend change are not too noisy then stock returns may have predictability in that variables that help predict future returns may exist. This would not invalidate market efficiency. The question is whether the predictability has economic value in the sense of generating abnormal returns. From this perspective standard measures of forecasting accuracy have modest information content, since they do not allow for transactions costs or map into the nature of the profit decision. For example a forecast that the market would rise by 12% when it rises by

5% is inferior from the perspective of a squared error measure of forecast accuracy to a forecast of 0%. Clearly a decision rule of the buy-sell type would value the forecast differently (see e.g. Pesaran, 1992; Granger and Pesaran, 1999). It is not clear as yet whether the predictability in asset returns generates abnormal returns particularly when the potential for changing risk is allowed for.

Strong-Form Market Efficiency

Insider trading in the major asset markets is subject to legal restrictions in a number of countries and has been illegal for a number of years. For instance in the US the Securities and Exchange Act of 1934 prohibits agents from trading securities while in possession of material inside information. Insiders are defined to include not only corporate insiders but also anyone who obtains material, non-public information from a corporate insider, or from the issuer, or who steals such information from another source. In subsequent years further acts were passed and in 1988 insider trading sanctions were further increased so that infringement can result in fines of up to one million dollars or five to ten years in jail. In the UK insider trading is a criminal offense and can result in up to two years in jail. As a consequence of this, indirect methods have to be employed to ascertain whether insiders can make abnormal returns on the basis of their private information. That insider trading does take place is *prima facie* supported by studies such as Keown and Pinkerton(1981) who found that on average 40-50% of the price gain experienced by a target firm's stock occurs before the actual takeover announcement.

Insiders who obey the law are in a number of countries, such as the US and UK, required to notify the stock exchanges of their 'routine trades' in the shares of their companies. This information is published reasonably quickly. A number of researchers have investigated whether this legal trading by corporate insiders can predict future stock market returns – e.g. Finnerty (1986), Friederich et al. (2000), Gregory et al. (1994), Pope et al. (1990), Seyhun (1986, 1992).These tests are semi-strong tests, since the information set underpinning the tests is public. The evidence is suggestive that excess returns can be obtained though the measurement of normal returns is an issue in some studies.

Jeng et al. (1999) examine the returns to corporate insiders themselves (based on legal trades) using a comprehensive sample of reported insider transactions from 1975 to1996. Their carefully-conducted study provides evidence of significant abnormal returns.

Meulbroek (1992) investigates the returns to illegal trades by insiders. Her sample consisted of individuals charged with insider trading during

1980-1989. Her data base included information on the charges brought, penalties incurred, profits earned, number of securities traded, type and source of the inside information. Among the defendants who traded, the median defendant transacted in one security and reaped \$17,628 in profit per security. By analysing security prices on the days insiders traded and did not prior to public announcements Meulbroeck could investigate the impact of insiders on stock price movements. She concluded that insider trading increases stock price accuracy by moving stock prices significantly. She found that the abnormal price movement on insider trading days is 40-50% of the subsequent price reaction to the public announcement of the inside information. It would appear therefore from the studies conducted so far that strong-form efficiency is not empirically supported.

The arguments for and against the legality of insider trading are discussed in e.g. Ausubel (1990), Benabou and Laroque (1992), Leland (1992), Khanna, Slezak, and Bradley (1994), Tighe and Michener (1994),

We now consider in more detail some tests for semi-strong market efficiency.

Variance Bounds Tests

An important test of the efficiency of present value models of asset prices was originally developed by LeRoy and Porter (1981) and Shiller (1979, 1981). This exploits the property of rational expectations that the variance of the outcome is greater than the variance of the rational expectation forecast.

For a variable Y_t

$$\text{var}(Y_t) = \text{var}(E_t Y_{t+i}) + \text{var}(v) \tag{141}$$

where v_t is the forecast error. This follows since rational expectations implies that the covariance between the forecast and the forecast error is zero.

In models of asset prices rational expectations implies that the actual variance of an asset price, P_t, which is a weighted average of future expectations of fundamentals should be less than the variance of the asset price computed on the basis of the actual stochastic process of the fundamentals.

For simplicity we consider the model in level rather than log level form. The asset price has the value, as we have seen in e.g. (33), of:

$$P_t = E_t \left[\sum_{i=1}^{\infty} \frac{D_{t+i}}{(1+\overline{R})^i} \right] \tag{142}$$

where D_{t+i} are the fundamental returns and $1+\overline{R}$ is the constant discount rate.

Assuming rational expectations we can decompose the present value relationship into two components one of which is the perfect foresight path and the other the sequence of rational expectations prediction errors:

$$P_t = \sum_{i=1}^{\infty} \frac{D_{t+i}}{(1+\overline{R})^i} - \sum_{i=1}^{\infty} \frac{u_{t+i}}{(1+\overline{R})^i} \tag{143}$$

where the u_{t+i} are the forecast errors. Consequently

$$P_t^* = P_t + v_t \tag{144}$$

where

$$P_t^* = \sum_{i=1}^{\infty} \frac{D_{t+i}}{(1+\overline{R})^i} \text{ and } v_t = \sum_{i=1}^{\infty} \frac{u_{t+i}}{(1+\overline{R})^i}$$

From (144)

$$\text{var}(P_t^*) = \text{var}(P_t) + \text{var}(v_t) \tag{145}$$

Hence plainly var $P_t \leq$var P_t^*.

The striking finding of the early literature was that stock prices appear to move too much to be consistent with subsequent changes in actual fundamentals. For example Shiller (1981, b) employing annual data on price and dividends from 1871 to1979 for the US stock market, so that the actual dividend stream is employed up to 1979 plus a proxy for 1980 to infinity (typically the actual last value of the asset price) to compute the variance in 1871, found that the actual asset price was some five times higher than the perfect foresight asset price. Leroy and Porter (1981) obtained similar results employing measures of earnings as fundamentals.

Two statistical problems with this analysis were pointed out by Flavin (1983), Kleidon (1986) and Marsh and Merton (1986). Computation of the sample variances in (144) involves use of the sample mean. In small samples Flavin and Kleidon demonstrate that use of the sample rather than population mean leads to a bias towards rejecting efficiency.

Marsh and Merton demonstrate that the variance bounds tests proposed by Shiller are not appropriate if the process generating the fundamentals are non stationary.

Mankiw, Romer and Shapiro (1985) develop a variance bounds test which is robust to these two points. Consider a naive forecast of the asset price, P_t^n

$$P_t^n = E_t^n \sum_{i=1}^{\infty} \frac{D_{t+i}}{(1+\overline{R})^i} \tag{146}$$

where E_t^n is the naive expectation of future fundamentals (say a constant growth of dividends model).

We can write

$$P_t^* - P_t^n = (P_t^* - P_t) + (P_t - P_t^n) \tag{147}$$

Taking the square of the left- and right-hand sides of (147) and taking unconditional expectations of t information

$$E(P_t^* - P_t^n)^2 = E(P_t^* - P_t)^2 + E(P_t - P_t^n)^2 \tag{148}$$

since by rational expectations $E(P_t^* - P_t)(P_t - P_t^n) = 0$ (i.e. the rational expectations forecast error, the first term in braces, is orthogonal to the second term)

Equation (141) implies that

$$E(P_t^* - P_t^n)^2 \geq E(P_t^* - P_t)^2 \tag{149}$$

and

$$E(P_t^* - P_t^n)^2 \geq E(P_t - P_t^n)^2 \tag{150}$$

The first expression states that the perfect foresight path is better forecast by the actual asset price than the naive forecast; the second expression states that the perfect foresight path fluctuates more around the naive path than does the actual price.

On the basis of tests that exploit these bounds Marsh and Merton also reject efficiency though the violation is smaller than in the original tests. In defense of the efficiency hypothesis we should note, as pointed out by various authors (e.g. LeRoy and LaCivita, 1981), that the models assume a constant discount rate or constant realized returns if the model is cast in loglinear form (see e.g. Campbell, Lo and Mackinlay, 1997), and the assumption that agents are risk neutral.

Empirical Tests: The Foreign Exchange Market

The foreign exchange market in particular has been looked at particularly exhaustively and is clearly of great interest for macroeconomics. We examine some of the empirical issues in detail in this section since many of the issues are relevant to other empirical work. A building block of many analyses is that covered interest arbitrage holds. Since this is an arbitrage condition, violation of it would be highly damaging to the efficient markets hypothesis. Numerous empirical studies have examined the covered condition. It appears that, once appropriate allowance is made for transactions costs and care is taken to ensure that data in empirical tests is sampled at the same moment in time, the condition is violated very infrequently (see e.g. Taylor, 1987, 1989). From the uncovered and covered arbitrage conditions we obtain that the logarithm of the forward exchange rate is equal to the market expectation of the logarithm of the future spot rate defined by the maturity date of the interest rate. In the case of a one period horizon we obtain

$$F_t^t = E_t S_{t+1} \tag{151}$$

Under the assumption of rational expectations we can rewrite (151) as

$$S_{t+1} = F_t^t + u_{t+1} \tag{152}$$

where u_{t+1} is the rational expectations forecast error so that $u_{t+1} = S_{t+1} - E_t S_{t+1}$.

Since S_{t+1}, and F_t^t are observable, early tests of the rationality of F_t^t were conducted by estimating the relationship

$$S_{t+1} = \alpha_0 + \alpha_1 F_t^t + \epsilon_{t+1} \tag{153}$$

and testing whether $\hat{\alpha}_0 = 0$, $\hat{\alpha}_1 = 1$ and the error term exhibited serial correlation.

These tests were supportive in general of the efficiency of the forward rate as a predictor. However it was realized subsequently that because the spot and forward rate appear to be nonstationary I(1) processes (Time-Series Annex), a coefficient of unity obtained in estimates of α_1 would be a weak test, even if appropriate standard errors were employed, since non-rational predictors which shared a common trend would exhibit this property.

Fama (1984) reported regression estimates for numerous currencies in the post-war floating period of the form

$$S_{t+1} - S_t = \gamma_0 + \gamma_1 (F_t^t - S_t) + \epsilon_{t+1} \tag{154}$$

Comparison of (152) and (154) shows that if the estimates of $\gamma_1$and γ_0 are not significantly different from one and zero respectively then the forward rate is an efficient predictor.

In fact Fama obtained the striking result that the estimates of γ_1 were negative and significantly different from unity (or zero). Qualitatively the same results have been obtained for numerous different data sets and time periods.

For example using monthly data from 1978.01 to 1990.07 McCallum(1994) reports the following results for the Dollar/Pound

$$S_{t+1} = -0.0137 + 0.977 F_t \quad \overline{R}^2 = 0.96 \tag{155}$$

$$S_{t+1} - S_t = -0.0078 - 4.7403(F_t^t - S_t) \ \ \overline{R}^2 = 0.111 \tag{156}$$

Whilst employing weekly data for the Dollar/Pound between December10th 1921–May 20th1925 we obtain

$$S_{t+4} = 0.121 + 0.920 F_t \ \ \overline{R}^2 = 0.86 \tag{157}$$

$$S_{t+4} - S_t = 0.0056 - 3.3166(F_t^t - S_t) \ \ \overline{R}^2 = 0.043 \tag{158}$$

The coefficients on the forward premium are significantly different from unity in these regressions and also significantly less than zero. The negative coefficient implies that not only is the forward rate an inefficient predictor, it is also inferior to the spot rate as a predictor of the future spot rate. These results appear to imply that the foreign exchange market is inefficient. Whilst this may be the case, and directly observed survey data of exchange rate expectations support such an interpretation since they typically exhibit bias and inefficiency (see e.g. Frankel and Froot, 1987, and Froot and Frankel, 1989) a number of possible resolutions of this apparent anomaly have been suggested which are consistent with market efficiency. We might also note in passing that if the outcome was due to inefficient expectations it poses a major problem for economic analysis given the systematic nature of the bias of the forward premium over 70 or so years.

The first explanation of the anomalous finding was set out by Fama (1984): the existence of a variable risk premium.

A time-varying risk premium

If we relax the assumption that investors are risk averse then agents will require a risk premium for undertaking the uncovered position. A variety of alternative theoretical derivations of the risk premium have been proposed (see e.g. Bekaert, 1996; Hansen and Hodrick, (1983), Hodrick, 1989; Sibert, 1989). In what follows we use the Consumption CAPM model.

The ex-post nominal return, $(1 + R_t^s)$, to an uncovered domestic transaction, is

$$1 + R_t^s = \frac{(1 + R_t^f) s_{t+1}}{s_t} \tag{159}$$

where R_t^f is the risk-free nominal foreign interest rate and s_t is the exchange rate measured as domestic currency per unit of foreign currency. Covered interest arbitrage for the risk-free nominal domestic interest rate, $(1+R_t)$ is given by

$$1+R_t = \frac{(1+R_t^f)f_t}{s_t} \tag{160}$$

where f_t is the forward exchange rate in levels. (Note that (68) is an approximation to (160) where $\log(1+R^f) \simeq R^f$, and $S = \ln s$)

Substitution out of the foreign interest rate from the covered condition and substitution in (159) gives

$$1+R_t^s = \frac{(1+R_t)s_{t+1}}{f_t} \tag{161}$$

The real ex-post domestic return from an uncovered speculation $(1+\overline{R}_t^s)$ is given by

$$1+\overline{R}_t^s = \frac{(1+R_t)s_{t+1}P_t}{f_tP_{t+1}} = \frac{(1+R_t)s_{t+1}s_tP_t}{s_tf_tP_{t+1}} \tag{162}$$

where P is the price level. The real *ex-post* domestic return from the risk-free nominal bond $(1+\overline{R}_t)$ is

$$1+\overline{R}_t = \frac{(1+R_t)P_t}{P_{t+1}} \tag{163}$$

From the consumption CAPM

$$\frac{E_t\beta(1+\overline{R}_t^s)u'(c_{t+1})}{u'(c_t)} = \frac{E_t\beta(1+\overline{R}_t)u'(c_{t+1})}{u'(c_t)} \tag{164}$$

Assuming conditional log normality and the form of utility function in (21) we can employ (164) and (162) to obtain

$$E_tS_{t+1} = F_t - 0.5\text{var}(\log\frac{s_{t+1}}{s_t}) + \text{cov}([\log s_{t+1}/s_t]\log P_{t+1}/P_t) + \delta\text{cov}([\log s_{t+1}/s_t]\log c_{t+1}/c_t) \tag{165}$$

(note again: $F = \log f,\ S = \log s$).

The risk premium reflects the covariation between changes in the logarithm of the exchange rate and the logarithm of changes in consumption. It is interesting to note that if $\delta = 0$, so that the consumer is risk neutral there are still terms driving a wedge between the expected future spot rate and the forward rate. These terms occur by virtue of what is known as Jensen's inequality.

If

$$f_t = E_ts_{t+1} \tag{166}$$

where we measure say in pounds to dollars we also require measuring in dollars to pounds that.

$$\frac{1}{f_t} = E_t \frac{1}{s_{t+1}} \tag{167}$$

However by Jensen's inequality

$$E_t\left(\frac{1}{s_{t+1}}\right) > \frac{1}{E_t s_{t+1}} \tag{168}$$

We cannot simultaneously have (166) and (167) holding which would appear to imply that one party can make expected nominal profits, known as Siegel's (1972) paradox. The resolution is that traders are interested in expected real profits. Given the assumption of lognormal distributions (165) shows the appropriate condition. The terms resulting from Jensen's inequality are generally presumed to be small but Sibert (1989) shows that this assumption may be inappropriate.

Addition of a risk premium to the uncovered arbitrage condition implies, in conjunction with the covered condition which remains unchanged (it is riskless), that

$$F_t = E_t S_{t+1} + p_t \tag{169}$$

where p_t is the 'risk premium' (inclusive of all the terms).

Subtracting the spot rate from both sides of (169) we obtain that

$$F_t - S_t = E_t S_{t+1} - S_t + p_t \tag{170}$$

Now consider the regressions

$$F_t - S_{t+1} = \alpha_1 + \beta_1(F_t - S_t) + \epsilon_{1t+1} \tag{171}$$

and

$$S_{t+1} - S_t = \alpha_2 + \beta_2(F_t - S_t) + \epsilon_{2t+1} \tag{172}$$

where the α and β are constants and the ϵ are the error terms.

The least squares estimates of β_1 and β_2 are given by

$$\beta_1 = \frac{\text{cov}(F_t - S_{t+1}, F_t - S_t)}{\text{var}(F_t - S_t)} = \frac{\sigma^2(p) + \text{cov}(p_t, E_t S_{t+1} - S_t)}{\sigma^2(p) + 2\text{cov}(p_t, E_t S_{t+1} - S_t) + \sigma^2(E_t S_{t+1} - S_t)} \tag{173}$$

and

$$\beta_2 = \frac{\text{cov}(S_{t+1} - S_t, F_t - S_t)}{\text{var}(F_t - S_t)} =$$

$$\frac{\sigma^2(E_tS_{t+1} - S_t) + \text{cov}(p_t, E_tS_{t+1} - S_t)}{\sigma^2(p) + 2\text{cov}(p_t, E_tS_{t+1} - S_t) + \sigma^2(E_tS_{t+1} - S_t)} \quad (174)$$

(recall in the least squares regression $Y = \alpha + \beta X$, the estimate of $\beta = \frac{\text{cov}(Y,X)}{\text{var}(X)}$) and that under rational expectations cov(ϵ_{t+1}, any variable at t) = 0. Note also that $S_{t+1} - S_t = E_tS_{t+1} - S_t + \epsilon_{t+1}$, where ϵ_{t+1} is the rational expectations forecast error.

We note the regressions are mirror images of each other with the property that $\beta_1 + \beta_2 = 1,\ \alpha_1 + \alpha_2 = 0$ and $\epsilon_{1t+1} + \epsilon_{2t+1} = 0$.

We note from the numerator of (174) that if the estimate of β_2 is negative this is consistent with a time-varying risk premium if the covariation between expected changes in the spot rate and the risk premium is negative and greater than the variance of expected changes in the spot rate.

Because β_1 is positive this implies that the variance of the risk premium is greater than the covariation between the expected change in the spot rate and the risk premium. Together these two conditions also imply that the risk premium must have a variance greater than the variance of expected changes in the spot rate. Hodrick and Srivatava (1986) showed how a negative covariation between expected changes in the spot rate and the risk premium could be obtained in a properly formulated theoretical model. Consequently the negative coefficient obtained in the regressions of changes in spot on the forward premium could be consistent with rational expectations and a time varying risk premium. Unfortunately data which would allow direct proxying of the risk premium are not available at the frequencies required. Time series methods based on modelling the risk premium as a function of the variance of the spot rate or forecast errors have not provided much direct support for a variable risk premium, though the models of the risk premia are possibly poor approximations – see e.g. Baillie and Bollerslev (1990) and Domowitz and Hakkio (1985), Bollerslev and Hodrick (1996) Engel (1996). In addition the magnitude of the variance of the risk premia implied by the estimates of β_2 are regarded by some as perhaps implausibly high (though see Pagan, 1988). For example an estimate of β_2 of -4 implies that $\text{var}(p_t) > 5\text{var}(F_t - S_t)$ (note $\beta_2 = -4 = 1 - \frac{(\text{cov}(p_t, E_tS_{t+1} - S_t) + \sigma^2(p_t)}{\text{var}(F_t - S_t)}$ is a different way of writing the estimate of β_2).

A second potential explanation of the empirical findings is called the 'peso problem', to which we now turn.

Peso Problems

The expression 'peso problem' was probably first employed by Milton Friedman in his analysis of the behaviour of the Mexican currency, the peso, in the early 1970s. Although the Mexican exchange rate was fixed between April 1954 and August 1976 at 0.08 dollars per peso, Mexican interest rates were substantially above US interest rates. This finding presented an apparent puzzle, since rational expectations of a continuing fixed exchange rate regime would, given negligible risk premia, imply via uncovered interest arbitrage near equality of Mexican and US interest rates of similar type. Friedman rationalized the differential in terms of a market expectation of a devaluation of the peso. In fact in August 1976 these expectations were subsequently realized when the peso was allowed to float and fell some 46% in value. The peso problem thus refers to a situation where rational agents anticipate the possibility of future changes in the data-generating mechanism of economic variables.

To illustrate the nature of the peso problem suppose there are two possible economic regimes in which the variable y_t follows the process

$$y_t = \overline{y}_1 + \theta y_{t-1} + u_t \text{ in regime 1} \tag{175}$$

or

$$y_t = \overline{y}_2 + u_t \text{ in regime 2} \tag{176}$$

where $\overline{y}_1$, $\overline{y}_2$ and θ are constants and u_t is a random variable which is assumed to be the same, for simplicity, in each regime.

Consider the rational expectation of y_{t+1} formed on the basis of the information set Ω available at time t, $E(y_{t+1} \mid \Omega_t)$. We assume that rational agents do not know which regime will prevail in the next period; rather they have probabilities q_t that regime 1 will occur and $1 - q_t$ that regime 2 will occur.

The rational expectation of y_{t+1} is then given by

$$E(y_{t+1} \mid \Omega_t) = q_t(\overline{y}_1 + \theta y_t) + (1 - q_t)\overline{y}_2 \tag{177}$$

The peso problem in its starkest form can be illustrated by assuming that only one of the regimes is ever observed in a sample of data. Consider the case where only regime 1 occurs. The rational expectations forecast error for this case is given by

$$\overline{y}_1 + \theta y_t + u_{t+1} - E(y_{t+1} \mid \Omega_t) = u_{t+1} + (1 - q_t)[\overline{y}_1 - \overline{y}_2] + (1 - q_t)\theta y_t \tag{178}$$

We observe from (178) that the *ex-post* forecast error in this case will exhibit bias and exhibit correlation with variables in the information set when expectations are formed (y_t). In fact if $\theta = 1$, in our example the forecast error will be non-stationary, a possibility pointed out by Evans and Lewis (1993, 1994).

It is also interesting to note what happens when a standard orthogonality or efficiency test for rational expections is run on data generated by this outcome, assuming q_t is constant: let the researcher estimate the relationship

$$y_{t+1} - E(y_{t+1} \mid \Omega_t) = \alpha + \beta(E(y_{t+1} \mid \Omega_t) - y_t) + \epsilon_{t+1} \tag{179}$$

where α and β are constants and ϵ_{t+1} is the error term.

The least squares estimate of $\widehat{\beta}$ =cov(Y, X)/var(X) where Y is the dependent variable and X the independent) is given by

$$\widehat{\beta} = \frac{-\theta(1-q)}{(1-\theta q)} \tag{180}$$

Consequently for $\theta > 0$ the regression estimate of β, which is negative, would imply that the rational expectation of y_{t+1} was an inferior predictor to the current level of y_t.

Non-occurrence of a regime (event) in a particular sample provides the extreme example of the peso problem. Such possiblities could occur when for instance there is a small probability of a change in regime. Such cases would not be amenable to differentiating empirically between rational expectations and an alternative unless the probability q and the alternative regimes could be described.

Evans (1996) has set out a general method of analysis of the peso problem some of which we now set out. We can define the rational expectations forecast error, e_{t+1}, as

$$e_{t+1} \equiv y_{t+1} - E(y_{t+1} \mid \Omega_t) \tag{181}$$

To examine how the properties of the forecast error are affected by the presence of discrete changes in regime Evans supposes that y_{t+1} can switch between two processes which are indicated by changes in a discrete-valued variable, $Z_t = \{0,\ 1\}$, so that Z_t only takes the value zero or unity. Let $y_{t+1}(z)$ denote realized returns in regime $Z_{t+1} = z$. The peso problem is to consider the behaviour of forecast errors, $y_{t+1} - E(y_{t+1} \mid \Omega_t)$. Evans does this by decomposing realized returns into the conditionally expected return in regime z, which is denoted $E(y_{t+1}(z) \mid \Omega_t)$ and a residual u_{t+1} (which for simplicity is assumed to be the same in both regimes).

The decomposition has the following form:

$$y_{t+1} = E(y_{t+1}(0) \mid \Omega_t) + \Delta E(y_{t+1} \mid \Omega_t) Z_{t+1} + u_{t+1} \tag{182}$$

where

$$\Delta E(y_{t+1} \mid \Omega_t) \equiv E(y_{t+1}(1) \mid \Omega_t) - E(y_{t+1}(0) \mid \Omega_t) \tag{183}$$

Substitution of (183) into (182) and setting Z_{t+1} as either one or zero informs us, as pointed out by Evans, that it will always be possible to decompose y_{t+1} in this way regardless of the process y_{t+1} follows in either regime or the precise specification of the information set Ω_t. Given the assumption of rational expectations $E(u_{t+1} \mid \Omega_t) = 0$, so that the error term u_{t+1} has the conventional rational expectations properties, Evans defines this error as the within-regime forecast error, since it represents the error when the $t+1$ regime is known.

When agents are unaware of the regime in $t+1$ their forecast errors will differ from the within-regime errors. Taking expectations of both sides of (182) on the basis of Ω_t information, we have from the properties of rational expectations that

$$E(y_{t+1} \mid \Omega_t) = E(y_{t+1}(0) \mid \Omega_t) + \Delta E(y_{t+1} \mid \Omega_t) E(Z_{t+1} \mid \Omega_t) \qquad (184)$$

where we note that the right-hand side now contains the expected regime in $t+1$.

Subtraction of (184) from (182) and substitution in (181) gives the *ex-post* forecast error as

$$e_{t+1} = u_{t+1} + \Delta E(y_{t+1} \mid \Omega_t)[Z_{t+1} - E(Z_{t+1} \mid \Omega_t)] \qquad (185)$$

We observe from (185) that when the regime at $t+1$ is known, so that $Z_{t+1} - E(Z_{t+1} \mid \Omega_t) = 0$, the forecast error is the standard within-regime error so that there is no peso problem. When the future regime is unknown the second term in (185) adds a component to the within-regime error (given that the within-regime forecasts differ so that $\Delta E(y_{t+1} \mid \Omega_t) \neq 0$).

Evans illustrates this point more clearly by supposing that in period $t+1$ regime 1 occurs. From (185) with $Z_{t+1} = 1$ we obtain that

$$e_{t+1}(1) = u_{t+1} + \Delta E(y_{t+1} \mid \Omega_t)[1 - E(Z_{t+1} \mid \Omega_t)] \qquad (186)$$

or

$$e_{t+1}(1) = u_{t+1} + \Delta E(y_{t+1} \mid \Omega_t) \Pr\{(Z_{t+1} = 0 \mid \Omega_t)\} \qquad (187)$$

where Pr denotes probability.

When the probability of regime 0 is non-zero the rational expectation error contains an additional component which is the difference between the within-regime forecasts multiplied by the probability that regime 0 occurs. As we illustrated with our example above when the within-regime forecasts differ this can generate an ex-post rational errors which may have a non-zero mean, so that they appear to be biased or serially correlated, and so inefficient. It is also apparent from (185) that the extent to which this issue will be important in empirical evaluations of forecasts will be dependent upon the frequency of regime shifts in the data sample. In the extreme case when only one regime occurs the errors will match those in (186). When there are a number of regime changes the forecast error will be a weighted average of $e_{t+1}(1)$ and $e_{t+1}(0)$ (defined anologously to (186)).The effect in this case on the properties of the forecast error will depend on the sample properties of the forecast error for regimes $Z_{t+1} - E(Z_{t+1} \mid \Omega_t)$. When the number of regime changes in the sample is representative of the underlying distribution of regime changes, from which the rational expectations of agents are generated, then the forecast error for regimes will exhibit a zero mean and the forecast errors will exhibit the standard rational expectations errors.

An interesting extension of the peso issue considered by Evans and referred to by Kaminsky (1993) as the generalised peso problem is when agents do not know which current regime they are in (for instance the central bank's preferences). Evans illustrates some implications by assuming that the degree of uncertainty of the regime is given by $\Pr(Z_t \mid \Omega_t)$. When the regime is known and there is no uncertainty $Z_t = z$ and $\Pr(Z_t \mid \Omega_t) = 1$. When $\Pr(Z_t \mid \Omega_t) \neq 1$ Evans discusses some implications in the following manner. First we employ the identity

$$\begin{aligned}\Pr(Z_{t+1} = 0) \equiv \Pr(Z_{t+1} = 0 \mid Z_t = 1) - \Pr(Z_{t+1} = 0 \mid Z_t = 1)\\ + \Pr(Z_{t+1} = 0) \end{aligned} \quad (188)$$

where for simplicity we have dropped the notation for the information set Ω_t. (We read e.g. the second term of (188), $\Pr(Z_{t+1} = 0 \mid Z_t = 1)$, as follows: the probability of $Z_{t+1} = 0$ given $Z_t = 1$, all conditional on the information set Ω_t).

We substitute (188) into (187) to obtain

$$\begin{aligned}e_{t+1}(1) = u_{t+1} + \Delta E(y_{t+1}) \Pr(Z_{t+1} = 0 \mid Z_t = 1) -\\ \Pr(Z_{t+1} = 0 \mid Z_t = 1) + \Pr(Z_{t+1} = 0)] \end{aligned} \quad (189)$$

or

$$\begin{aligned}e_{t+1}(1) = u_{t+1} + \Delta E(y_{t+1})\{\Pr(Z_{t+1} = 0 \mid Z_t = 1\} -\\ \Delta E(y_{t+1})[\Pr(Z_{t+1} = 0 \mid Z_t = 1) - \Pr(Z_{t+1} = 0)] \end{aligned} \quad (190)$$

Now

$$[\Pr(Z_{t+1} = 0 \mid Z_t = 1) - \Pr(Z_{t+1} = 0)] \quad (191)$$

the last term in (190), is equal to

$$[\Pr(Z_{t+1} = 0 \mid Z_t = 1) - \Pr(Z_{t+1} = 0 \mid Z_t = 0)] \Pr(Z_t = 0) \quad (192)$$

since equating (191) and (192) and rearranging gives the definitional statement

$$\Pr(Z_{t+1} = 0) = \Pr(Z_{t+1} = 0 \mid Z_t = 1)[1 - \Pr(Z_t = 0)] + \Pr(Z_{t+1} = 0 \mid Z_t = 0)[\Pr(Z_t = 0)] \quad (193)$$

(note $1 - \Pr(Z_t = 0) = \Pr(Z_t = 1)$).

Consequently the last term in (190) can be rewritten as

$$-\Delta E(y_{t+1})[\Pr(Z_{t+1} = 0 \mid Z_t = 1) - \Pr(Z_{t+1} = 0 \mid Z_t = 0)] \Pr(Z_t = 0) \quad (194)$$

Recognizing that the first two terms on the right-hand side of (190) are the same as those in (187) we see that uncertainty about the current regime manifests itself in an additional term in the forecast error given by (194). We observe that this term will only be zero if the probability of regime 0 at time $t + 1$ is independent of the current regime or of course that within-regime forecasts are equal. In addition we observe that changes in $\Pr(Z_t = 0)$ (perhaps due to learning about the regime) will contribute to the structure of the error term.

Evans (1996) documents a number of other interesting implications of the peso problem. In the context of asset prices, where current values depend on expected values of fundamentals into the indefinite future, he demonstrates how news about future fundamentals can influence current prices through the normal channel of changed forecasts of future fundamentals as well as the additional channels of the difference in fundamentals in the different regimes as well as the dynamics of regime switching. In the context of empirical work he demonstrates how peso problems can contribute to the explanation of some of the puzzles found in asset markets, such as the bias of the forward premium. Bekaert, Hodrick and Marshall (2001) show how peso problems can contribute to an explanation of anomalies in the term structure of interest rate (see below).

The peso problem is that sample statistics are not representative of the population so that statistical inferences are potentially misleading. It is clearly of considerable importance in evaluating models of expectations.

Statistical Rationale

An alternative statistical rationale for the forward premium bias is suggested by Baillie and Bollerslev (1997). An empirical regularity in the exchange market is that the statistical properties of changes in the spot rate and the forward premium are markedly different. In monthly data the variance of changes in the spot rate is typically some 100 or so times greater than the variance of the forward premium. In addition whilst changes in the spot rate typically exhibit little evidence of serial correlation and can often be well approximated by martingales the forward premium typically displays very persistent slowly decaying autocorrelations which may be described as a fractional process – see e.g. Baillie and Bollerslev (1994), Byers and Peel (1996), also the Time-Series Annex). The statistical properties of changes in the spot rate and the forward premium imply that estimates of β_2 in small samples may be fragile. Baillie and Bollerslev show this is the case. They report estimates of β_2 in the Fama regression obtained from a rolling sample of data. Employing monthly data for the DM/dollar they consider 208 five-year rolling regression estimates for β_2 obtained by beginning in March 1973 and using a total of sixty observations through to February 1978. Then the next estimate was obtained by using data from April 1973 through to March 1978, until the final estimate was based on data from December 1990 through to November 1995. The estimates of β_2 display marked instability varying between -13 and 3.52. These findings suggest that for many sample sizes encountered in practice, the estimate of β_2 is likely to be uninformative about the true value of β_2.

They provide further support for this view by simulating data from a known structure which is calibrated to produce the stylized statistical properties exhibited by the spot and forward premiums. The expectation embodied in the forward premium is rational by construction. Their simulated models are found to generate results which are similar to those reported in the literature. In particular, the forward premium exhibits persistent autocorrelation and the estimates of β_2 are widely dispersed between negative and positive numbers.These empirical findings are of some interest since the statistical properties exhibited by the dependent and independent variables in the Fama regression are qualitatively similar to those exhibited in other tests of market efficiency such as bond rates (see below).

McCallum (1994) illustrates another important point that needs to be given consideration when testing market efficiency relationships. This is whether the relationship is invariant to government policy. To illustrate, McCallum assumes that the monetary authorities manage interest rates so as to smooth their movements while also resisting changes in exchange rates.

Defining $x_t = R_t - R_t^F$ he considers the policy rule

$$x_t = \lambda(S_t - S_{t-1}) + \sigma x_{t-1} + e_t \tag{195}$$

where $\lambda,\ \sigma$ are constants and e_t is a random error.

Uncovered arbitrage is given by

$$x_t = E_t S_{t+1} - S_t + v_t \tag{196}$$

McCallum interprets v_t as measurement error though it could be interpreted as a risk premium. He assumes that

$$v_t = \rho v_{t-1} + u_t \tag{197}$$

where ρ is a constant and u_t is a random error. Assuming a solution for changes in the spot rate of the form

$$S_t - S_{t-1} = Ax_{t-1} + Bu_t + Ce_t \tag{198}$$

and equating coefficients we obtain the solution as

$$S_t - S_{t-1} = \frac{(\rho - \sigma)}{\lambda} x_{t-1} + \frac{u_t}{(\lambda + \sigma - \rho)} - \frac{e_t}{\lambda} \tag{199}$$

(note that $x_{t-1} = R_{t-1} - R^f_{t-1} = F_{t-1} - S_{t-1}$). The solution for x_t can also be obtained as

$$x_t = \rho x_{t-1} + \frac{\lambda u_t}{(\lambda + \sigma - \rho)} \tag{200}$$

The important point to note is that the coefficient on x_{t-1} ($= F_{t-1} - S_{t-1}$ via the covered arbitrage condition) in (199) contains the policy parameters λ and σ. If σ is close to unity and greater than ρ and λ is positive, then the coefficient can be negative. In addition, the forward premium will exhibit persistent serial correlation for ρ close to unity and if the variance of u_t is very small relative to the variance of e_t then changes in the spot rate will be closely approximated by a white noise error process even though the true process is a persistent ARMA(1, 1) process.

This concludes our discussion of tests of market efficiency in the exchange market. It has emphasised the considerable difficulties in determining from *ex-post* regression whether this market is efficient.

We next consider, more briefly, the implications of market efficiency for the behaviour of interest rates.

LONG-RUN INTEREST RATES UNDER MARKET EFFICIENCY

There are a great variety of bonds. Bonds are issued by private companies, local authorities and by the government. Naturally the probability of default can vary across bonds with different issuers. But government bonds are normally regarded by investors as free of default risk and it is the pricing of these types of bonds in an efficient market that we will consider here. Government bonds can be broadly categorized as being of two main types. The first type are bonds which pay a regular coupon, usually semi-annual, which is a fixed fraction of the face value of the bond up to the maturity date when the face value is also repaid. The other type are zero-coupon bonds, also called discount bonds which make one payment at the maturity date of the bond. Bonds are offered with a great variety of maturity dates, from short up to undated (consols in the UK). Bonds paying coupons can be interpreted as packages of zero-coupon bonds with one corresponding to each coupon payment and one corresponding to the repayment of the principal and coupon at the maturity date. For an individual trader, long-run bonds are substitutable for short-run bonds, since it is possible for him to hold a series of short bonds rather than a long bond over the same holding period, or conversely to hold a long bond for a short period and then sell it rather than hold a short bond to maturity.

More formally, assume initially that traders are completely certain of the future. Initially we consider pure discount bonds so that the return is simply the discount from par at which the bond is sold at the beginning of the period relative to the redeemed par price at the end of the period.

The price, P_{nt}, at time t of a bond which makes a payment of £1 at time $t+n$ is given by

$$P_{nt} = \frac{1}{(1+R_{nt})^n} \tag{201}$$

where R_{nt} is the bond's yield to maturity.

In a world of certainty the pure expectations hypothesis assumes that the the following condition must hold:

$$(1+R_{nt})^n = (1+R_{1t})(1+R_{1t+1})(1+R_{1t+2})\ldots(1+R_{1t+n-1}) \tag{202}$$

The left hand side of (202) is simply the rate of return on holding an n-period bond until maturity. The right side of (202) is the rate of return implied by holding a one-period bond for one period, then reinvesting the proceeds (principal plus interest) in a one-period bond for the next period and so on.

By taking logarithms and recalling that for small values of the fraction Z, $\ln(1+Z)$ can be approximated by Z, we can rewrite (202) as:

$$R_{nt} = \frac{1}{n}(R_{1t} + R_{1t+1} + R_{1t+2} + \ldots + R_{1t+n-1}) \tag{203}$$

In other words (203) informs us that long-run interest rates are simply averages of future interest rates over the time period to maturity. It tells us, for instance, that if short-run interest rates remain constant for the indefinite future, then long rates will be equal to short rates. Conversely, if future short rates are expected to fall, the current long rate will be below the short rate. The relationship is known as the expectations theory of the term structure of interest rates.

When we relax the assumption of perfect knowledge of the future and recognize that traders can observe the current long rate R_{nt} and the current one-period short rate R_{1t}, it follows if traders are risk neutral that:

$$R_{nt} = \frac{1}{n}[R_{1t} + E_t R_{1t+1} + E_t R_{1t+2} + \ldots E_t R_{1t+n-1}] \tag{204}$$

If traders are risk averse, then we must add a risk premium to the right-hand side of (204) but this does not affect the argument, provided it is constant.

If we consider a bond with a two-period maturity date then

$$R_{2t} = \frac{1}{2}[R_{1t} + E_t R_{1t+1}] \tag{205}$$

We can rewrite (205) as

$$E_t R_{1t+1} - R_{1t} = 2(R_{2t} - R_{1t}) \tag{206}$$

or assuming rational expectations that

$$R_{1t+1} - R_{1t} = 2(R_{2t} - R_{1t}) + v_{t+1} \tag{207}$$

where $v_{t+1}$is the expectations forecast error. Accordingly we observe that the expected change in yield on a one-period bond is related to twice the difference between the yield on a two-period bond and the yield on a one-period bond and that the actual change differs from this term by an expectational error. The expectations hypothesis implies testable implications such as this for bonds of different maturities. We now examine some of these implications further.

We first introduce the concept of the holding-period return, H_{nt+1}, which for simplicity we assume is one-period. This is the return from purchasing an n-period bond at time t and selling it at time $t+1$. At time $t+1$ the bond will become an $n-1$ period bond and will be sold for a price P_{n-1t+1} .

From (201)

$$1 + H_{nt+1} = \frac{P_{n-1t+1}}{P_{nt}} = \frac{(1+R_{nt})^n}{(1+R_{n-1t+1})^{n-1}} \tag{208}$$

The expectations hypothesis also implies that

$$R_{1t} = E_t[H_{nt+1}] \tag{209}$$

Equation (209) tells us that the yield on a one-period bond should equal, ignoring any risk premium, the expected holding-period return on an n-period bond held for one period. Taking logs of (208) and approximating as above, and then taking expectations, we can use (209) to eliminate the expected holding-period yield. Manipulation of the resultant gives:

$$E_t(R_{n-1t+1} - R_{nt}) = \frac{(R_{nt} - R_{1t})}{n-1} \tag{210}$$

Equation (210) informs us that when the spread (the difference in yield between the long-maturity and short-maturity bond) is positive, changes in yields on long-maturity bonds are expected to increase. Campbell, Lo and Mackinlay (1997) and others have tested (210) by estimating the equation

$$R_{n-1t+1} - R_{nt} = \alpha + \beta\frac{(R_{nt} - R_{1t})}{n-1} + \varepsilon_t \tag{211}$$

where ε_t is the forecast error.

The estimate of β should not differ from unity when equation (211) is estimated for discount bonds of different maturity dates. The results are not supportive of the expectations hypothesis. Estimates of β are often negative, particularly for bonds of long maturity dates.

A variety of explanations exist for this anomalous finding and they mimic the explanations for the anomolous empirical finding in the exchange market discussed above whereby the forward premium is negatively related to future changes in the spot rates. These are time-varying risk premia, peso problems, goverment policy rules and low statistical power of the regressions.

We can also write (204) in the form

$$R_{nt} - R_{1t} = \frac{1}{n}[R_{1t} + (E_tR_{1t+1} - R_{1t} + R_{1t}) + (E_tR_{1t+2} - R_{1t} + R_{1t}) + \ldots + (E_tR_{1t+n-1} - R_{1t} + R_{1t})] - R_{1t} \tag{212}$$

or

$$R_{nt} - R_{1t} = E_t\sum_{i=1}^{n}\frac{\Delta R_{1t+i-1}}{n} \tag{213}$$

Equation (213) informs us that the spread is a predictor of future changes in short run interest rates. This implication can also be empirically tested. One method is to replace the expected terms on the right hand side of (213) by their actual values and regress the resultant on the spread, employing an estimator that corrects for the overlapping expectational errors. The coefficient should not differ from unity. Campbell, Lo and Mackinlay (1997) report such estimates. Their results suggest a U-shaped pattern in the estimates for different maturities. For short maturities, the spread has a positive coefficient which is less than unity and declines initially, becoming insignificant with the maturity horizon. At longer horizons a significant positive coefficient is obtained sometimes greater than unity. It thus appears that the spread has ability to predict short-run interest rates changes for both relatively short and long horizons but not intermediate horizons. The empirical results can again be explained by reference to the arguments explaining previous anomalous findings above.

We now turn to analysis of coupon-paying bonds.

The yield to maturity, or long-run interest rate Y_t^n, on an n-period bond is determined by the fact that the price P_t^n of the bond is the present value of a coupon (C), assumed paid at the end of each period, and repayment of the principal (normalized here to unity) at the terminal date discounted by Y_t^n. We have that

$$P_t^n = \frac{C}{1+Y_t^n} + \frac{C}{(1+Y_t^n)^2} + \frac{C}{(1+Y_t^n)^3} + \ldots + \frac{C}{(1+Y_t^n)^n} + \frac{1}{(1+Y_t^n)^n} \quad (214)$$

Letting $u = \frac{1}{1+Y_t^n}$ we can multiply (214) by u to obtain

$$uP_t^n = Cu^2 + Cu^3 + Cu^4 + \ldots + Cu^{n+1} + u^{n+1} \quad (215)$$

Subtracting (215) from (214) we obtain

$$P_t^n(1-u) = Cu - Cu^{n+1} + u^n - u^{n+1} \quad (216)$$

Dividing (216) by $1-u$, noting that $P_t^n(1-u) = Cu(1-u^n)+u^n(1-u)$, and $\frac{u}{1-u} = \frac{1}{Y_t^n}$, rearranging after substitution back for u we obtain that

$$P_t^n = \frac{C}{Y_t^n} + \frac{Y_t^n - C}{Y_t^n(1+Y_t^n)^n} \quad (217)$$

If $Y_t^n = C$, $P_t^n = 1$, and conversely. Bonds with this characteristic, whose price today equals the principal paid at maturity are selling ‘at par’.

We also notice from (217) that for an undated security or perpetuity, $n = \infty$ and so $P_t = \frac{C}{Y_t}$, so that the price is the coupon divided by the very long-term interest rate, Y_t.

The one-period holding yield, H_t^n, on a n-period coupon-paying bond consists of both the yield in the holding period plus the capital gain or loss. This gives us

$$H_t^n = \frac{P_{t+1}^{n-1} - P_t^n + C}{P_t^n} \tag{218}$$

recalling again that an n-period bond at time t becomes an $n-1$ bond at time $t+1$.

We note from (218) that for a perpetuity, $n = \infty$,

$$H_t = Y_t - \frac{(Y_{t+1} - Y_t)}{Y_{t+1}} \tag{219}$$

Equating the expected holding-period yield from the perpetuity to the yield on a short-term coupon-paying bond we have that

$$Y_t^1 = Y_t - \frac{(Y_{t+1} - Y_t)}{Y_{t+1}} \tag{220}$$

If we take a first-order Taylor expansion of the last term on the right-hand side of (220) around $\overline{Y}$ we obtain

$$Y_t^1 = Y_t - 1 + 1 + \frac{(Y_t - \overline{Y})}{\overline{Y}} - \frac{\overline{Y}(Y_{t+1} - \overline{Y})}{\overline{Y}^2} \tag{221}$$

Rearranging (221) we obtain in expectational form

$$E_t Y_{t+1} - Y_t = \overline{Y}(Y_t - Y_t^1) \tag{222}$$

We observe that expected changes in the yield on the perpetuity are positively related to the spread between the yield on the perpetuity and the yield on the short maturity bond.

Equation (222) can also be solved forward to obtain the solution for Y_t as

$$Y_t = \frac{\overline{Y}}{1 + \overline{Y}} E_t \sum_{i=0}^{\infty} \frac{Y_{t+i}^1}{(1 + \overline{Y})^i} \tag{223}$$

so that the perpetuity is a weighted average of expected future short yields. Clearly (223) is a linear approximation to a nonlinear structural equation and its ability to approximate will naturally be dependent on interest rates not being too volatile so that the approximation remains reasonably accurate

In the more general case of coupon-paying bonds with an n-period horizon the derivation can proceed in a similar manner. We substitute for the price of the bond from (217) into the holding period yield (218). This gives us the rather messy relation

$$H_t^n = \frac{\left(C + \frac{C}{Y_{t+1}^{n-1}} + \frac{Y_{t+1}^{n-1} - C}{Y_{t+1}^{n-1}(1+Y_{t+1}^{n-1})^{n-1}}\right)}{\left(\frac{C}{Y_t^n} + \frac{Y_t^n - C}{Y_t^n(1+Y_t^n)^n}\right)} - 1 \qquad (224)$$

(In Shiller, 1979, equation (224) is linearized around $Y_t^n = Y_{t+1}^{n-1} = \overline{Y} = C$.) This gives the relationship

$$H_t^n = \frac{Y_t^n - \gamma_n Y_{t+1}^{n-1}}{1 - \gamma_n} \qquad (225)$$

where $\gamma_n = \{1 + \overline{Y}[1 - 1/(1+\overline{Y})^{n-1}]^{-1}\}^{-1}$

Equating (225) to a short yield, Y_t^1 (plus any constant risk premium) we can solve the resulting difference equation for Y_t^n, given the terminal condition that the price of the bond is 1 at $t+n$, as

$$Y_t^n = \frac{1-\gamma}{1-\gamma^n} E_t \sum_{i=0}^{\infty} \gamma^i Y_{t+i}^1 + \phi_n \qquad (226)$$

where ϕ_n is any constant risk or liquidity premium. Equation (226) informs us that the n-period yield on a coupon-paying bond is a weighted average of the expected yields on the one-period bond.

Shiller exploits these properties to evaluate empirically whether long-run bond yields are too volatile to be consistent with rational expectations and the observed volatility in short rates (a test analogous to the variance bounds test for stock prices outlined above). His empirical findings are inconsistent with rational expectations. Again the results could be reconciled by appeal to the sort of issues raised above.

LEARNING AS AN ALTERNATIVE TO RATIONAL EXPECTATIONS:

The tests we have been examining all assume rational expectations. One possible cause of failure could therefore be that expectations are not rational but rather the result of a learning procedure — of which adaptive expectations is an approximation, as shown by Benjamin Friedman (1979).

We observed in the appendix to chapter 2 that if a series followed an ARIMA(0,1,1) process then an adaptive expectations scheme could in fact be a representation of the rational expectation. Earlier in this chapter we also found that a regressive expectations scheme could be rational in the Dornbusch overshooting model. In general of course these methods of expectations formation will not have the rational expectations property. They illustrate however that in any particular model structure, since the rational expectation solution can be written in terms of observable variables, there is a 'mechanistic' method of expectation formulation which will have the properties of rational expectations for the particular model structure. In this book we are primarily concerned with investigating the properties of models when it is assumed that agents' expectations, or the aggregate of agents' expectations, have the rational expectations property. Maintained assumptions of this approach are that agents know the true model, or act as though they did, and also assume that other agents also possess this information (see Townsend, 1978). In this approach econometricians will come across cases of 'adaptive' expectations which by chance represent the rational forecasting mechanism.

However there is an important literature that relaxes these assumptions and assumes that expectations follow a learning rule,which has the potential to converge to rational expectations (see e.g. Evans and Honkapohja, 1999, 2001, and Sargent, 1993). From this perspective the adaptive or regressive expectations schemes can be interpreted as simple learning rules which have converged on the rational expectations solution. A variety of learning rules, in general possibly more intelligent, have been studied. One such example is least-squares learning whereby agents employ least-squares regression to estimate the parameters of a model and employ the resulting model to forecast the variable of interest. This approach to learning models agents in the same way as economists who employ econometrics and statistical inference in deciding between competing models. Employing this approach also highlights another aspect of rational expectations that agents in the model are assumed to possess more information than the outside observer.

One motivation for studying learning rules is to ascertain whether they converge on rational expectations and how fast this process is. Answers to such questions may be relevant for some in deciding how plausible the rational expectations assumption is.

Whilst the implausibility of the assumption that agents know the true model may be sufficient justification for some for studying the implications of the learning assumption there are other properties of the rational expectations assumption that have been used to justify such an approach.

The first of these is the issue of 'non-uniqueness' (as set out in chapter 2). When the model exhibits more than one stationary solution methods that are not part of the formal model structure have to be employed to determine which solution is chosen by agents. The solution chosen by different learning rules has been widely studied. However, as we argued in chapter 2, 'non-uniqueness' should really be seen as the failure of a stability condition for the 'forward' root. The ultimate remedy is to find a specification which satisfies stability conditions that presumably hold in the real world.

The second justification is when a model has more than one equilibrium solution. For example, take the widely used model of hyperinflation which has been extensively studied in this context. If we assume a government prints money to finance a constant budget deficit, then

$$P_t G_t = M_t - M_{t-1} \tag{227}$$

where P_t is the price level, $G_t = \overline{G}$ is the constant real deficit, and M_t is the money stock.

Assume a demand function for money of the form

$$\frac{M_t}{P_t} = f(E_t p_{t+1}) \tag{228}$$

where $E_t p_{t+1} = E_t(\log(P_{t+1}/P_t)$ is the expected rate of inflation:

$$\frac{\partial \frac{M_t}{P_t}}{\partial E_t p_{t+1}} < 0 \tag{229}$$

and real output has been assumed constant.

Assuming money demand is equal to money supplied we can substitute (2) into (1) to obtain

$$\overline{G} = f(E_t p_{t+1}) - f(E_{t-1} p_t) e^{-p_t} \tag{230}$$

(since $\log \frac{P_t}{P_{t-1}} = p_t, \frac{P_t}{P_{t-1}} = e^{p_t}$)

Equation (3) has two equilibrium solutions if $\overline{G}$ is not too large (where $E_t p_{t+1} = E_{t-1} p_t = p_t = \overline{p}$). Intuitively if inflation is zero in equilibrium the authorities generate no receipts, whilst if inflation is infinite agents hold no money so receipts are also zero. Consequently the equilibrium surface $\overline{G} = g(\overline{p})$ exhibits two equilibria for $\overline{G} < \overline{G}_{\max}$. Assuming rational expectations the equilibrium exhibiting higher inflation is mathematically stable and the lower one unstable. These rankings are reversed assuming adaptive expectations. If the view is taken that mathematical stability is not the appropriate selection criteria in a rational expectations model then there is no mechanism, which is part of the formal model structure, to choose between the two equilibria. More generally a rational expectations models could exhibit multiple stable equilibria.

A third justification is structural change. Say a new government or a new central banker appears, it is natural to model agents as learning about the new regime. How 'wet' or 'dry' the new governor is and how the probabilities change as new information accrues seem best analyzed in a learning framework.

We will now briefly illustrate some examples of learning mechanisms. Evans and Honkapohja (1999) have divided the approaches into three groups and we follow their taxonomy.

Eductive approaches

In this literature researchers investigate whether the coordination of expectations on an rational expectations equilibrium can be attained by a mental process of reasoning. We illustrate with an example based on DeCanio(1979).

Suppose the demand and supply in a market are given by

$$q_t = a - bp_t + w_t \tag{231}$$

$$q_t = c + dE_{t-1}p_t + v_t \tag{232}$$

where p_t is the price level, w_t and v_t are random disturbances and a, b, c and d are constants.

The reduced form for prices in this system assuming demand is equal to supply is given by

$$p_t = \frac{a-c}{b} - \frac{d}{b}E_{t-1}p_t + \frac{w_t - v_t}{b} = A - BE_{t-1}p_t + u_t \tag{233}$$

where the definitions of A,B,u_t are obvious.

Suppose agents form their expectations initially in an arbitrary manner.The question is whether they can modify their behaviou r in such a way as to lead them closer to rational expectations, given by $\frac{A}{1+B}$. Suppose the initial, arbitrary expectation of all agents is given by

$$E^0_{t-1}p_t = p_{t-1} \tag{234}$$

From (232) given this expectation the actual evolution of prices will be given by

$$p_t = A - Bp_{t-1} + u_t \tag{235}$$

DeCanio assumes that after some passage of time agents realize (reason or deduce) that prices are evolving according to (234) and form the new expectation

$$E^1_{t-1}p_t = A - Bp_{t-1} \tag{236}$$

However this new expectation changes the evolution of the system to

$$p_t = A - B(A - Bp_{t-1}) + u_t = A - BA + B^2 p_{t-1} + u_t \tag{237}$$

Agents observing the new evolution of prices in the market agents revise expectations to

$$E^2_{t-1} p_t = A - BA + B^2 p_{t-1} \tag{238}$$

so that actual prices evolve as

$$p_t = A - B(A - BA + B^2 p_{t-1}) + u_t = A - BA + B^2 A - B^3 p_{t-1} + u_t \tag{239}$$

Continuing in this manner after n iterations we will have

$$E^n_{t-1} p_t = A - BA + B^2 A - B^3 A + - + AB^n + B^n p_{t-1} \tag{240}$$

If $|B| < 1$, for large n expectations will converge to the rational expectation

$$E^n_{t-1} p_t = \frac{A}{1+B} \tag{241}$$

since $\frac{1}{1+B} = 1 - B + B^2 - B^3 + -+$ for $|B| < 1$.

Clearly convergence to the rational expectation is not guaranteed even in the simple example if $|B| > 1$. When iterative expectations converge on the rational expectations solution the rational expectation is said to be iteratively E-stable. The iterative expectations of agents were assumed to be homogenous in the above example. When convergence occurs and the iterative expectations of agents are heterogeneous as in Guesnerie (1992) the rational expectations model is said to be strongly rational. Evans (1985, 1986) employ the iterative expectations method in models embodying multiple solutions, Peel and Chappell (1986) in a model embodying multiple equilibria and Bullard and Mitra (2000) in a model where agents learn about monetary policy rules.

Adaptive Approaches

Early on Benjamin Friedman (1979) argued that agents would learn from data via regression about the model and the policy regime. This, he pointed out, would produce expectations formation very similar to adaptive expectations without necessarily ever leading to rational expectations. Such statistical learning was subsequently examined to see whether it would converge on rational expectations. We consider the least-squares learning mechanism initially analyzed by Bray and Savin (1986) and Fourgeaud, Gourieroux and Pradel (1986), though we note that more complicated estimation procedures such as neural nets and genetic algorithms have been employed (see Sargent, 1993).

Suppose for simplicity that the reduced form for prices follows the process

$$p_t = A + BE_{t-1}p_t + Cz_{t-1} + u_t \tag{242}$$

From (241) the rational expectation is given by $E_{t-1}p_t = \frac{A+Cz_{t-1}}{1-B}$ so that prices evolve as

$$p_t = A + \frac{(AB + C)}{1-B}z_{t-1} + u_t = A + Gz_{t-1} + u_t \tag{243}$$

Suppose agents believe that prices follow the process given by (242) but are unaware of the values of the parameters A and G. In the least-squares approach to learning agents are assumed to run least-squares regressions of p_t on z_{t-1} and an intercept using previous data on the variables. Expectations are then generated from the estimated model. As more data becomes available the model is then reestimated, expectations formed and so on. Researchers have demonstrated that the conditions for convergence of recursive least-squares expectations can be weaker than those under iterative expectations. ($B < 1$ in this model as opposed to $|B| < 1$ with iterative expectations.)

In the case of least-squares learning where agents perceive the reduced form as

$$y_t = \beta' x_t + e_t \tag{244}$$

where β' is a vector of coefficients, x_t a vector of explanatory variables and e_t an error, the least-squares estimated coefficients are given by

$$\beta_t = \left(\sum_{i=0}^{t-1} x_i x_i'\right)^{-1} \left(\sum_{i=0}^{t-1} x_i y_i\right) \tag{245}$$

It can be demonstrated that the recursive least-squares estimates are generated as

$$\beta_t = \beta_{t-1} + \gamma_t R_t^{-1} x_{t-1}(y_{t-1} - \beta_{t-1}' x_{t-1}) \tag{246}$$

and

$$R_t = R_{t-1} + \gamma_t (x_{t-1}x_{t-1}' - R_{t-1}) \tag{247}$$

with $\gamma_t = \frac{1}{t}$. and where R_t is an estimate of the moment matrix for x_t. For suitable initial conditions $R_t = t^{-1} \sum_{i=0}^{t-1} x_i x_i'$.

We note that the term $\gamma_t = \frac{1}{t}$ is known as the 'gain'. It is important in determining the speed of convergence to the true parameter. Some intuition of the least-squares updating formulae can be obtained by considering the recursive least-squares estimate of the mean $Ez_t = \mu$. The least-squares estimate is the sample mean $\overline{z}_t = \frac{1}{t} \sum_{n=1}^{t} z_n$.

Subtracting the sample mean at $t-1$ from both sides of $\overline{z}_t$ and rearranging gives

$$\overline{z}_t = \overline{z}_{t-1} + \frac{1}{t}(z_t - \overline{z}_{t-1}) \tag{248}$$

since $t\overline{z}_t = \sum_{n=1}^{t} z_n = z_t + \sum_{n=1}^{t-1} z_n$ and $(t-1)\overline{z}_{t-1} = \sum_{n=1}^{t-1} z_n$ so that $t(\overline{z}_t - \overline{z}_{t-1}) = z_t + \sum_{n=1}^{t-1} z_n - t\overline{z}_{t-1}$
$= z_t + (t-1)\overline{z}_{t-1} - t\overline{z}_{t-1} = z_t - \overline{z}_{t-1}$.

Adaptive methods of learning have the same general type of structure which is given by

$$\theta_t = \theta_{t-1} + \lambda_t Q(t, \theta_{t-1}, X_t) \tag{249}$$

where θ_t is a vector of parameters, λ_t is the gain parameter equal to $\frac{1}{t}$ in the case of least-squares, Q is a function and X_t is the vector of variables in the structural model. We note that adaptive expectations is a special case of (248) where the gain parameter is constant..

The evolution of X_t will depend on θ_{t-1}, in the case of a linear system

$$X_t = A(\theta_{t-1})X_{t-1} + B(\theta_{t-1})W_t \tag{250}$$

where W_t is a vector of disturbance terms.

Marcet and Sargent (1989a,b), Evans and Honkapohja (1998) derive stability results for linear (and nonlinear) systems. Sargent (1999) assumes the US authorities used constant-gain least-squares learning about the Phillips Curve and maximised a social objective function to pick inflation; he argues this fits US post-war data, accounting for the 'great inflation' where rational expectations cannot.

Rational Learning

Rational learning has to be interpreted from a perspective that acknowledges the benefits and costs of more accurate forecasts for an agent as in Feige and Pierce (1976) or Evans and Ramsey (1992), so that rational expectations may not be attained unless calculation costs are zero. However the method most widely employed to model rational learning has been to employ Bayes' rule. Bayes' rule is the basic property of conditional probability and is a method of updating our belief or probability of event or hypothesis A given new evidence B. Specifically, our posterior belief $P(A/B)$ is calculated by multiplying our prior belief $P(A)$ by the likelihood $P(B/A)$ that B will occur if A is true. To see this we can rearrange the conditional probability formula to get:

$$P(A/B)P(B) = P(A, B) \tag{251}$$

where $P(A, B)$ is the joint probability of A and B.

By symmetry we also have

$$P(B/A)P(A) = P(A, B) \tag{252}$$

It follows from (250) and (251) that .

$$P(A/B) = \frac{P(B/A).P(A)}{P(B)} \tag{253}$$

Equation (252) is called Bayes' rule or Bayes' theorem. The formulation carries the implication that beliefs change by learning: agents come to know of a new fact and form their posterior belief by conditioning their prior belief on these facts.

An alternative form of Bayes' rule is given by

$$P(A/B) = \frac{P(B/A).P(A)}{\sum\limits_i P(B/A_i).P(A_i)} \tag{254}$$

since $P(B) = \sum\limits_i P(B/A_i).P(A_i)$ where A_i refers to the event space.

From (252) we observe that the data or new facts B only influence the posterior inference through the function or probability $P(B/A)$ which is called the likelihood function.

The ratio of the posterior probability evaluated at the two events A_1 and A_2 is called the posterior odds.

We have

$$\frac{P(A_1/B)}{P(A_2/B)} = \frac{P(A_1)P(B/A_1)/P(B)}{P(A_2)P(B/A_2)/P(B)} = \frac{P(A_1)P(B/A_1)}{P(A_2)P(B/A_2)} \tag{255}$$

In words, the posterior odds are equal to the prior odds multiplied by the likelihood ratio.

Bayes' rule has been widely employed to model learning in the economics literature — see e.g. Cyert and DeGroot (1974), Backus and Driffill (1985), Ellison and Valla (2000), Lewis (1998) Sill and Wrase (1999), Townsend (1978). Types of problem analyzed employing Bayes'rule include learning about a new regime. We will give one example which is a slight modification of Lewis (1988). She demonstrates how beliefs that a policy process may have switched can induce apparent ex-post biased forecasts of exchange rates even after the switch has occurred. In addition, in her model, exchange rates may appear to contain a speculative bubble component since they will systematically deviate from the levels implied by observing fundamentals ex-post.

Assume the reduced form for the exchange rate (see chapter 14) is given by

$$s_t = m_t + \alpha(E_t s_{t+1} - s_t) \tag{256}$$

where s_t is the money supply at time t, s_t is the exchange rate and α is a positive constant.

Assume the money supply process is given by

$$m_t = \theta_0 + \varepsilon_t^0 \tag{257}$$

where θ_0 is a constant and ε_t^0 is a normally distributed random variable with mean zero and variance σ_0^2.

Suppose at a particular point in time, say $t = 0$, agents come to believe that the money supply process may have changed, due to an exogenous process such as a change in government or a statement by officials. The new process has for simplicity the same form as the old except with a different mean and variance:

$$m_t = \theta_1 + \varepsilon_t^1 \text{ for t} \geq 0 \tag{258}$$

It is assumed that $\theta_1 < \theta_0$ and $\theta_1 = 0$, so that the process can be interpreted as going from 'loose' to 'tight' money. Agents are not sure which money supply process is in operation. It is further assumed for simplicity that agents believe that if policy has changed it will not be changed back and they also know the parameters of the potential new process.

Solving (29) forward we obtain the solution

$$s_t = (1-\gamma)\sum_{i=0}^{\infty}\gamma^i E_t m_{t+i} \tag{259}$$

where $\gamma = \frac{1}{1+\alpha}$.

Expected money supply given the assumptions above is equal to

$$E_t m_{t+i} = \theta_0(1 - P_{1t}) \text{ for any } i > 0, t \geq 0 \tag{260}$$

where P_{1t} is agents' assessed probability at time t that the process changed at time 0.

Consequently the exchange rate is given by

$$s_t = (1-\gamma)m_t + \gamma(1-P_{1t})\theta_0 \tag{261}$$

(note $E_t s_{t+1} = \theta_0(1-P_{1t})$ from (30) and (31) since $\frac{1}{1-\gamma} = 1+\gamma^1+\gamma^2+ -+\gamma^\infty$ and $\frac{1}{1+\alpha} = 1-\gamma$)

To obtain the best estimate of P_{1t}, agents combine their prior beliefs about the probability together with their observations of money outcomes each period to update their posterior probabilities according to Bayes' Rule.

$$P_{1t} = \frac{P_{1t-1}f(I_t/\theta_1)}{P_{1t-1}f(I_t/\theta_1) + P_{0t-1}f(I_t/\theta_0)} \tag{262}$$

where P_{0t} is the probability of no change at $t = 0$, $f(I_t/\theta_{_i})$ is the probability of observing the information set I_t given that m_t follows the ith process. The posterior probabilities of each process, the posterior odds ratio, is given by

$$\frac{P_{1t}}{P_{0t}} = \left[\frac{P_{1t-1}f(m_t/\theta_1)}{P_{0t-1}f(m_t/\theta_0)}\right] = \left[\frac{P_{1t-1}}{P_{0t-1}}\right]\left[\frac{(\frac{1}{\sigma_1})e^{(\frac{-m^2}{2\sigma_1^2})}}{(\frac{1}{\sigma_0})e^{(\frac{(m-\theta_0)^2}{2\sigma_0^2})}}\right] \tag{263}$$

The first term on the right-hand side of (262) shows that the change from $t-1$ to t in the relative conditional probabilities depends upon the observation of the current money supply at time t. For example for some observation of current money supply, say $\overline{m}$, the probability of being under either regime is equal, so that the posterior probabilities,$\frac{P_{1t}}{P_{0t}}$, are equal to the prior probabilities$\frac{P_{1t-1}}{P_{0t-1}}$. Observations of money supply different from $\overline{m}$ convey information about the regimes causing the probabilities to be revised. The last term on the right-hand side of (262) quantifies this information. For a normally distributed error the sampling distribution for a single scalar observation, m, from a normal distribution parameterized by a mean θ and a variance σ^2 is given by

$$P(m/\theta) = \frac{e^{-\frac{(m-\theta)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \tag{264}$$

Clearly the information accruing from an observation will depend on the particular distribution of the error term.

If for simplicity we assume that $\sigma_0 = \sigma_1 = \sigma$ then we can write (36) as

$$\ln(\frac{P_{1t}}{P_{0t}}) = \ln\frac{P_{1t-1}}{P_{0t-1}} + \ln\frac{f(m_t/\theta_1)}{f(m_t/\theta_0)} \tag{265}$$

so that for any realization of the money process m_k

$$\ln \frac{f(m_k/\theta_1)}{f(m_k/\theta_0)} = \left[\frac{(m_k - \theta_0)^2 - m_k^2}{2\sigma^2} \right] \tag{266}$$

Substituting (265) into (266) we obtain a linear difference equation that can be solved as

$$\ln(\frac{P_{1t}}{P_{0t}}) = \ln \frac{P_{1,0}}{P_{0,0}} + \sum_{k=1}^{t} \left[\frac{(m_k - \theta_0)^2 - m_k^2}{2\sigma^2} \right] \tag{267}$$

In equation (266) the evolution of probabilities is seen to depend on the actual observations of the process. When the money supply observed today is very negative agents will think it more likely that policy has changed and vice versa.

Taking expectations of (266), defining θ_i as the true θ gives

$$E \ln(\frac{P_{1t}}{P_{0t}}) = \ln \frac{P_{1,0}}{P_{0,0}} + t \left[\frac{(\theta_i - \theta_0)^2 - \theta_i^2}{2\sigma^2} \right] \tag{268}$$

$$= \ln \frac{P_{1,0}}{P_{0,0}} + t \left[\frac{\theta_0^2 - 2\theta_i\theta_0}{2\sigma^2} \right] \tag{269}$$

Equation (268) illustrates that the expected value of the 'true' process rises over time. For instance when policy has changed so $\theta_i = \theta_1 = 0$, from (268) we observe that the log probability increases to infinity because of the trend term $\frac{t\theta_0^2}{2\sigma^2}$ (or that P_{0t} goes to zero). When policy has not changed so that $\theta_i = \theta_o$ the reverse is true due to the trend term $\frac{-t\theta_0^2}{2\sigma^2}$. Consequently agents' assessed probabilities of a policy change are random variables determined by random observations of the money supply.

Given this analysis of the evolution of probabilities Lewis is able to investigate the effects on the exchange rate and forecast errors.Taking expectations at $t-1$ of (260) and subtracting from (260) we obtain the forecast errors corresponding to each potential process as:

$$s_t - E_{t-1}s_t = (1-\gamma)\varepsilon_t^0 + \theta_0(P_{1,t-1} - \gamma P_{1t}) \ \text{ if } \ \theta_i = \theta_0 \tag{270}$$

and

$$s_t - E_{t-1}s_t = (1-\gamma)\varepsilon_t^1 - \theta_0(P_{0,t-1} - \gamma P_{0t}) \ \text{ if } \theta_i = \theta_1 \tag{271}$$

(note $E_{t-1}P_{1t} = P_{1,t-1}, E_{t-1}P_{0t} = P_{0,t-1}$).

Equations (269) and (270) illustrate that whilst agents are learning, expectation errors will exhibit a systematic component and 'appear non-rational *ex-post*'.

Taking expectations of the forecast errors in (270) conditional upon a change in policy to θ_1 the expected evolution of the forecast errors, for a large number of m_k is given by

$$E(s_t - E_{t-1}s_t/\theta_1) = -\theta_0[E(P_{0,t-1}/\theta_1) - \gamma E(P_{0t}/\theta_1)] < 0 \qquad (272)$$

The inequality is negative since γ is less than one and $EP_{0t}/\theta_1 < E(P_{0,t-1}/\theta_1)$.

From (271) we observe that if agents do not fully realize that policy has changed the exchange will be expected to be weaker than subsequently occurs. Ex-post expectations will appear to be irrational.

Lewis' model nicely illustrates how learning about a regime change using Bayes' rule can mimic the outcomes of the Peso problem discussed earlier in this chapter.

CONCLUSIONS

In this chapter we have discussed some of the implications of the proposition that financial markets are efficient. As particular applications of the proposition we examined the behaviour of stocks, exchange rates and bond markets. The theoretical work on efficiency in asset markets is large and is expanding dramatically so that we have only been able to provide a flavour of the debate. Whilst the new generation of models which stress micro-structure arguments appear of great interest, our own view is that the empirical evidence for (approximate) semi-strong efficiency in the capital market is sufficiently powerful and convincing for it to be regarded as a 'stylized fact'. In fact it is now standard practice for macroeconomic model builders to simulate or forecast the impact of changes in government policy within models that assume capital market efficiency.

APPENDIX 14A1 – INCOMPLETE CURRENT INFORMATION

The purpose of this appendix is to show the manner in which the standard test of efficiency based on equation (81) of this chapter has to be modified if agents have an incomplete current information set. For analytical simplicity (though the argument is quite general), we assume the series y_t is the summation of two infinite moving average error processes in the two white noise errors, ϵ and z.

$$y_t = \overline{y} + \sum_{i=0}^{\infty} \pi_i \epsilon_{t-i} + \sum_{i=0}^{\infty} \delta_i z_{t-i} \tag{1}$$

.where $\overline{y}$ is the mean of the series, and the π_i and δ_i are constant coefficients.

Consider the rational expectation of y_{t+1} formed at time t. If there is full current information at time t the expectation will be given by:

$$E_t y_{t+1} = \overline{y} + \sum_{i=0}^{\infty} \pi_i \epsilon_{t-i+1} + \sum_{i=0}^{\infty} \delta_i z_{t-1+i} \tag{2}$$

In this case, given the one-period forecast horizon, the ex-post forecast error will be given by a white noise error:

$$y_{t+1} - E_t y_{t+1} = \pi_0 \epsilon_{t+1} + \delta_0 z_{t+1} \tag{3}$$

Consequently the standard tests based on (81) are correct in these circumstances.

Suppose next that agents have incomplete current information at time t, and instead observe some current global information (for instance via asset markets), but other global information with a one-period lag. In particular we will assume for simplicity (though the argument is easily generalized) that there is one global indicator (say the interest rate) which is given the representation:

$$R_t = \overline{r} + \sum_{i=0}^{\infty} (d_i \epsilon_{t-i} + h_i z_{t-i}) \tag{4}$$

where $\overline{r}$ is the mean of the series and the d_i and h_i are constant coefficients.

In this incomplete current information case the one-period-ahead expectation of y_t is given by:

$$E_t y_{t+1} = \overline{y} + \pi_1 E_t \epsilon_t + \sum_{i=2}^{\infty} \pi_i \epsilon_{t-i+1} + \delta_1 E_t z_t + \sum_{i=2}^{\infty} \delta_i z_{t-i+1} \tag{5}$$

Consequently the forecast error is given by:

$$y_{t+1} - E_t y_{t+1} = \pi_0 \epsilon_{t+1} + \delta_0 z_{t+1} + \pi_1[\epsilon_t - E_t\epsilon_t] + \delta_1[z_t - E_t z_t] \quad (6)$$

Given current observation of the global indicator $R_{t,}$ and using the usual signal extraction formulae (as discussed in chapter 3) we obtain:

$$E_t\epsilon_t = \frac{1}{d_o}\phi_\epsilon(d_0\epsilon_t + h_0 z_t) \quad (7)$$

$$E_t z_t = \frac{1}{h_0}(1-\phi_\epsilon)(d_0\epsilon_t + h_0 z_t) \quad (8)$$

where

$$\phi_\epsilon = \frac{d_o^2\sigma_\epsilon^2}{d_o^2\sigma_\epsilon^2 + h_o^2\sigma_z^2}$$

and σ_ϵ^2 and σ_z^2 are the variances of the two errors, ϵ and z respectively. Consequently the forecast error (K_{t+1}) is given by:

$$K_{t+1} = \pi_0\epsilon_{t+1} + \delta_0 z_{t+1} + \pi_1\left[(1-\phi_\epsilon)\epsilon_t - \frac{h_o}{d_o}\phi_\epsilon z_t\right] + \delta_1\left[\phi_\epsilon z_t - \frac{d_0}{h_o}(1-\phi_\epsilon)\epsilon_t\right] \quad (9)$$

Serial Correlation of Forecast Errors

If we take expectations of two successive errors we find:

$$E(K_{t+1},\ K_t) = \left\{\frac{\sigma_z^2\sigma_\epsilon^2}{d_o^2\sigma_\epsilon^2 + h_0^2\sigma_z^2}\right\}[\pi_1 h_0 - d_1 d_0].[\pi_0 h_0 - \delta_0 d_0] \quad (10)$$

Consequently, in general, incomplete current information will give rise to a moving-average error process. This will not be the case if, first, we have implicitly full current information (for example, if there are as many global indicators in the economy as random shocks; see Karni, 1980), or if, secondly, we observe the current value of the variable to be forecast (it being itself a global indicator). In this latter case $\pi_0 = d_0,\ \delta_0 = h_0$ (also $\pi_1 = d_1,\ \delta_1 = h_1$) and the expected correlation in (10) is equal to zero.

It would appear from this result that standard tests of efficiency based on (81) will have the usual properties for asset prices in particular (which it can be assumed are observed currently), even under incomplete current information. However, the general point remains that variables not currently observed (i.e. the majority) will under incomplete information be inappropriately tested for efficiency by these methods. Furthermore, one needs to scrutinize carefully the assumption that the asset prices in question are contemporaneously observed. In very high frequency data (e.g. hourly) this will obviously not be so except for a few continuously-broadcast asset prices; it will also not be so in lower frequency data for averages of variables (e.g. the level of all short-term interest rates), which are often examined in these studies.

In general, in circumstances of incomplete current information the moving-average error in equation (9) will be given by $s + j - 1$, where s is the time horizon of the forecast and j is the longest lag on global information relevant for forecasting y_t. Clearly there may be some *a priori* doubt as to the magnitude of j, which may cause some problems in interpretation of tests based on (81). As a consequence of the moving-average error process, least squares estimates of (81) under incomplete current information will be inefficient but unbiased, since

$$E(K_{t+1},\ E(y_{t+1})) = 0 \tag{11}$$

and least squares estimates have the property of unbiasedness even in the presence of moving-average error processes. This situation is the same as that of overlapping information in the usually assumed case of full current information; overlapping information here occurs with $s > 1$, familiarly introducing a moving-average process with the same effects.

These results have potential implications for a number of empirical studies (see e.g. Holden and Peel, 1977; Turnovsky, 1970) in which an implicit assumption of full current information has been made when studying directly-observed consumer price expectations data which cannot readily be assumed to be part of the current information set. The point here is that price data are not currently observable on any reasonable assumptions. Consequently, the 'expectations errors' should be serially correlated, as indeed has often been found in these tests. It is possible that these survey data may well reveal rationality after all. For further implications of partial current information sets for the testing of efficiency, see Minford and Peel (1984).

APPENDIX 14A2 COMPOSITE MOVING-AVERAGE ERROR PROCESSES

Testing for the efficiency of relationships involving moving-average errors poses problems. We wish to form a moving-average error process from the composite error process

$$av_{t+1} + bv_t + cu_{t+1} + du_t \tag{1}$$

where v_{t+i} and $u_{t+i}(i = 1,\ 0)$ are serially uncorrelated random variables.

We define the new moving-average process

$$\phi_{t+1} + j\phi_t = av_{t+1} + bv_t + cu_{t+1} + du_t \tag{2}$$

where ϕ_{t+i} $(i = 1,\ 0)$ is a serially uncorrelated random variable.

The method is to equate the ratio of the variance to the covariance of the error processes on the left- and right-hand sides of (2) (where the variances and covariances of the left- and right-hand sides are equal by definition — a similar method is employed for higher-order composite processes).

We obtain that

$$\frac{(1+j^2)\sigma_\phi^2}{j\sigma_\phi^2} = \frac{(a^2+b^2)\sigma_v^2 + (c^2+d^2)\sigma_u^2 + 2(bd+ac)\mathrm{cov}(uv)}{ab\sigma_v^2 + cd\sigma_u^2 + (ad+bc)\mathrm{cov}(uv)} \tag{3}$$

where σ_ϕ^2, σ_v^2, σ_u^2 are the variances of ϕ, v, u and $\mathrm{cov}(uv)$ is the covariance between u and v.

Equation (3) is a quadratic equation in j (note σ_ϕ^2 cancels). The root of the equation which has modulus less than unity is chosen so that the process can be stationary. Clearly the magnitude of j will reflect relative magnitudes of variances and covariances. For example if $a = 1,\ b = -1,\ c = 1$ and $d = 0,\ \sigma_u^2 = \sigma_v^2$ and $cov(uv) = 0$ then $j = \frac{-3+5^{0.5}}{2}$.

Using the lag operator it is also useful to note that we can express ϕ_{t+1} as

$$\phi_{t+1} = \frac{[av_{t+1} + bv_t + cu_{t+1} + du_t]}{(1+jL)} = [av_{t+1} + bv_t + cu_{t+1} + du_t]\{1 - jL + j^2L^2 - j^3L^3 + ...] \tag{4}$$

From (4) we observe that ϕ_{t+1} can be expressed as an infinite sum of the past errors v and u. This is important in some tests of efficiency under rational expectations. Since the errors in the composite process are in principle any errors in the economy if a moving-average error process is estimated jointly with the other parameters of the model it will induce correlations between the regressors and the error term which will result in biased and inconsistent parameter estimates. For instance consider the two-period ahead rational expectations forecast error for the process,

$$y_t = \overline{y} + \sum_{i=0}^{\infty} \gamma_i u_{t-i} + \sum_{i=0}^{\infty} \delta_i v_{t-i} \tag{5}$$

We have that

$$y_t = E_{t-2}y_t + \gamma_0 u_t + \gamma_1 u_{t-1} + \delta_0 v_t + \delta_1 v_{t-1} \tag{6}$$

Although the error term is a moving-average process, in a test of efficiency it would be inappropriate to estimate jointly the parameters of the model and the moving-average process

$$y_t = \alpha_0 + \alpha_1 E_{t-2}y_t + \phi_t + j\phi_{t-1} \tag{7}$$

and test that $\hat{\alpha}_0 = 0$, $\hat{\alpha}_1 = 1$. As shown above the error term can be written as an infinite summation of previous errors. This is also a property of the forecast so that there will be correlation between the error and the explanatory variable. An appropriate procedure is to estimate (7) by least squares and employ standard errors which are modified to allow for the serial correlation in the error term (see e.g. Hansen and Hodrick, 1980).

It is also useful to note that an ARMA forecast of a variable, even assuming the underlying model is linear, is less efficient than a rational expectations forecast if the reduced form of a variable includes composite moving-average errors. Suppose for illustration a variable, y_t, is generated by the process

$$y_t = u_t - u_{t-1} + v_t \tag{8}$$

The rational one-period ahead forecast, $E_{t-1}y_t = -u_{t-1}$. Let the moving-average process for (8) be given by $y_t = \phi_t - j\phi_{t-1}$ $(j > 0)$ so that the one-period ARMA forecast is $E^a_{t-1}y_t = -j\phi_{t-1}$. The associated forecast errors are

$$y_t - E_{t-1}y_t = u_t + v_t \tag{9}$$

for the rational expectation and for the ARMA forecast

$$y_t - E^a_{t-1}y_t = \phi_t = \frac{u_t - u_{t-1} + v_t}{(1 - jL)} \tag{10}$$

Assuming for simplicity that $cov(uv)$ is zero the variance of the rational expectations forecast is

$$\sigma_u^2 + \sigma_v^2 \tag{11}$$

The ARMA forecast error can be written as

$$y_t - E^a_{t-1}y_t = u_t - (1-j)u_{t-1} - j(1-j)u_{t-2} - j^2(1-j)u_{t-3} + \dots \\ + v_t + jv_{t-1} + j^2 v_{t-2} + \dots \quad (12)$$

(Recall that $\frac{1}{(1-jL)} = 1 + jL + j^2L^2 + j^3L^3 + \dots$)

Although the ARMA forecast error appears serially correlated this is in fact not the case as substitution for j in terms of the variances of u and v, though messy, will demonstrate.

The variance of the ARMA forecast error is given by

$$\sigma_u^2 + \sigma_u^2(1-j)^2[1 + j^2 + j^4 + j^6 + \dots] + \\ \sigma_v^2\{1 + j^2 + j^4 + j^6 + \dots\} \quad (13)$$

which we can simplify as

$$\sigma_u^2 + \frac{(1-j)^2\sigma_u^2}{(1-j^2)} + \frac{\sigma_v^2}{(1-j^2)} \quad (14)$$

Because j is less than one the variance of the ARMA forecast is greater than the rational expectations forecast error. Essentially information is lost in forecasting the composite error process. In addition the innovation from the ARMA process, ϕ_t, though serially uncorrelated, will not necessarily be orthogonal to variables that are correlated with past u and v innovations, due to the implicit dependence of the ARMA innovation on these variables.