Rational Exuberance*

Stephen F. LeRoy
University of California, Santa Barbara
September 27, 2003

1 Introduction

Consider the postage stamp. As title to a future good (or, in this case, service) with monetary value, this humble object is essentially the same as a security. Its value, 37 cents, is the present value of the service (delivery of a letter) to which its owner is entitled.

Now consider a postage stamp with a minor printing error that excites stamp collectors’ interest. Because of the printing error the stamp has a value of $1000. Viewing the stamp as a security suggests two possible explanations for the difference between the values of the stamp with and without the printing error: (1) the fundamental value of the stamp with the printing error is higher than that without, and (2) the stamp with the printing error has a bubble (these terms are defined below).

Attributing the higher value of the stamp with the printing error to its greater fundamental value would be justified if one believes that stamp collectors derive pleasure from contemplating the printing error. Alternatively, the collector may acquire status of the requisite value in the eyes of other collectors based on his or her ownership of the misprinted stamp, and this is the basis for the higher value. Along these lines, fluctuations in asset prices are necessarily attributed to fluctuations in preferences. Such arguments are best seen as making an inference about utility based on reverse-engineering the price fluctuations: someone’s marginal utility must be changing, or why would the price fluctuate?

George J. Stigler and Gary S. Becker [86] argued persuasively against relying on assumed preference shifts to explain such price fluctuations, especially when there

---

*Apologies to Alan Greenspan, whose famous 1996 speech raised the possibility of “irrational exuberance” in the stock market, and to Robert J. Shiller, who developed the idea in his book [83] of the same name. I have received comments from Ted Bergstrom, Christian Gilles, John Griffin, Michael Magill, Martine Quinzii, Matt Spiegel, Douglas Steigerwald, Shyam Sunder and Jan Werner. Francisco Azeredo found several errors.
exist alternative explanations of price fluctuations that are well grounded in economic theory. In the present context, if assets have bubbles their prices do not necessarily have the close association with agents’ utilities that the above argument presumes. Specifically, in the presence of bubbles asset prices can exceed the discounted value of their payoffs or service flows, generally by a random amount. Therefore asset price fluctuations can be explained without resorting to random preferences.

Contrary to a widely-held opinion, asset price bubbles do not necessarily reflect a model specification that is exotic in any respect (unless assuming an infinite number of dates is considered exotic). Rather, bubbles are like Giffen goods: although somewhat counterintuitive, they can occur under circumstances that are not difficult to understand (as with Giffen goods, establishing theoretical conditions under which bubbles can occur is much easier than determining empirically whether they do occur).

It is easiest to argue against the proposition that the prices of collectibles reflect fundamentals when the objects in question clearly have minimal aesthetic value, as with misprinted postage stamps. With fine art the argument is less straightforward: someone who prefers Monet to Andy Warhol might argue that the values of Monet’s paintings reflect their aesthetic value, whereas those of Andy Warhol’s paintings reflect bubbles. Even here, however, the fact that a Monet forgery that is undetectable except to an expert—and therefore presumably has the same aesthetic value as an original to anyone who is not an expert—has negligible monetary value argues that the value of a Monet original is a bubble, just as with the misprinted postage stamp.

Even if bubbles exist on collectibles, there is no guarantee that they exist on assets like stocks or land, and still less that bubbles on such assets are quantitatively important if they do exist. It is possible that there are arguments against bubbles that apply to securities but not to collectibles. However, the reasoning just presented creates a presumption in favor of bubbles: if highly valued fine art objects are bubbles, why would the stock of corporations that own art objects not have a bubble component?

Ponzi schemes—pyramids in which the contributions of new investors are used to pay high returns to earlier investors—and chain letters are other possible examples of bubbles, although interpretation of Ponzi schemes is clouded by the fact that investors have difficulty distinguishing them from highly successful genuine investments until it is too late.

The number of papers developing the economic theory of asset price bubbles has grown exponentially since the last major survey paper was published (Colin Camerer [17]; see also the Journal of Economic Perspectives 1990 symposium). At least equally important in motivating another visit to this topic, asset prices in the US and elsewhere have undergone the gyrations that are widely—and correctly, if the argument of this paper is accepted—associated with bubbles. This is a good time to provide a review of the economic literature on bubbles, parts of which are technical, that is as simple as possible (but, in Einstein’s famous phrase, not more so). The
focus is on stock prices. Contrary to what appears to be majority opinion, we will find
that economic theory provides no strong presumption against bubbles. The same is
true of the empirical evidence which, for the purpose of this survey, consists of the fact
that stock prices and returns are more volatile than they would be in the absence of
bubbles. None of this, of course, establishes incontrovertibly the existence of bubbles.
As always, theoretical arguments depend ultimately on the plausibility of alternative
sets of assumptions, about which there is no consensus, and the empirical evidence
is amenable to various interpretations even more with bubbles than elsewhere in
economics.

It is true that there exist possible explanations for observed stock price volatility
other than bubbles. Of these the most attractive on theoretical grounds is stoch-
astic discount factors (of course, this is not to deny that bubbles and stochastic
discount factors can coexist). Risk aversion induces a stochastic component in dis-
count factors, thereby generally invalidating the constant discount factor version of
the present-value model used in the variance bounds tests. Specifically, risk aversion
induces asset price volatility greater than that which would occur if agents were risk
neutral, at least in some settings. However, the theory of consumption-based asset
pricing, while generating a theoretically coherent account of how risk aversion affects
asset prices, turns out to be a failure empirically.

The remainder of the paper deals with bubbles. Bubbles are usually taken to
be synonymous with irrationality. This characterization is universal in the popular
press, but is also sometimes seen in academic discussions. Especially given the
recent ascendancy of behavioral finance, it is appropriate to digress to consider what,
if anything, it means to appeal to irrationality in a substantive economic explanation
of economic phenomena such as asset price fluctuations.

The subsequent discussion is restricted to rational bubbles—instances of asset
prices exceeding present values in models in which agents optimize in an environment
that they understand. As noted, established professional opinion holds that there
exist compelling theoretical arguments against existence of such bubbles. It is true
that in an important class of models—overlapping generations models—bubbles can
be ruled out under certain values for key parameters, but this argument for nonex-
istence of bubbles requires an implausibly literal-minded application of a rational
expectations argument to events in the distant future.

In this paper the intent is to convince readers that an appeal to bubbles to ex-
plain the volatility of asset price fluctuations does not entail a radical departure from
received methods of economic analysis. Conceding this does not guarantee that
bubbles do in fact constitute the explanation, particularly for those who do not share
Stigler and Becker’s reluctance to appeal to taste shifts. Ultimately it is an empirical
question, and the empirical literature on bubbles is not yet well developed. Perhaps
readers will turn their attention to developing empirical tests that can reliably distin-
guish between bubbles and other phenomena that affect asset prices, of which there
is now a shortage.
2 The US Data

Figure 1 shows the value of US equity (from the Federal Reserve Board Flow of Funds accounts) divided by GDP. Data are from the 1950s to now. The salient feature of this series is its dominance by low-frequency components. Stock prices rose in the 1950s and 1960s, fell in the 1970s, rose in the 1980s and 1990s and, so far in the 21st century, have been falling. The high-frequency component of stock price variation that dominates financial reporting is seen to be of minor importance by comparison with the low-frequency variation just discussed. The runup over the period 1995-2000 is conspicuous, as is the even more rapid subsequent collapse in prices.

Figure 2 shows stock values normalized instead by National Income Accounts corporate earnings. For most of the sample period the two figures look similar. To the extent that price-earnings ratios show variations similar to price-GDP ratios, the interpretation is that stock price variations cannot be viewed as simply a proportional response to parallel trends in earnings. For most of the sample period the low-frequency variation in Figure 2 is positively correlated with, but less pronounced than, that in Figure 1. The interpretation is that stock price variations are correlated with variations in earnings, with the stock price variations being of greater amplitude.

\footnote{For historical discussion of bubbles, see Charles Mackay’s classic 1841 volume [64]. More recent treatments are Peter Garber [33] for the early bubbles in Europe and John Kenneth Galbraith [32] for the 1929-1932 US stock price selloff. General introductions were provided in Charles P. Kindleberger [53], Edward Chancellor [19] and Shiller [83].}
The runup of stock prices in the late 1990s appears similar in the two series, suggesting that it cannot be viewed as a proportional response to spectacular increases in earnings. Indeed, Figure 3, which displays aftertax corporate earnings as a proportion of GDP, shows that while earnings rose in the middle and late 1990s, even at their peak they were a smaller proportion of GDP than during most of the postwar period.

In contrast, the collapse in stock prices over the past several years that is conspicuous in Figure 1 has no counterpart in Figure 2. The reason is that, as Figure 3 shows, corporate earnings underwent an even more pronounced drop than stock prices as the US economy entered a recession. Consequently the price-earnings ratio rose even as the price-GDP ratio fell. Partial data for 2002 (not displayed), in fact, show a further increase in the price-earnings ratio.

Our first question is whether these variations in stock prices are within the range that is appropriate under the present-value model, given the variation of corporate earnings and dividends. In the next section we use a variance-bound test to show that they are not.

3 The Stochastic Gordon Model

The point of reference in interpreting stock price data will initially be the Gordon model (Myron J. Gordon [40]). Subsequently we will generalize to a stochastic version of the model.
3.1 The Deterministic Gordon Model

The deterministic version of the Gordon model is generated by five assumptions: (1) the return on invested capital is constant over time, (2) the rate of earnings retention (equivalently, dividend payout) is constant over time, (3) retained earnings generate the same returns as preexisting capital (so that there are no opportunities for extranormal earnings), (4) the value of equity equals the discounted value of future dividends (so that there are no bubbles), and (5) the discount factor is constant over time, with the discount rate equal to the rate of return on invested capital. Under these assumptions it is easy to show that the present-value model implies that $p$, the value of equity, is given by

$$ p = \frac{d(1 + g)}{r - g} = \frac{e(1 + g)}{r}. \tag{1} $$

Here $r$ is the discount rate, $d$ is the current dividend, $e$ is current earnings and $g = r(1 - \delta)$ is the growth rate of the dividend stream, where $\delta$ is the dividend payout rate (proportion of earnings paid out as dividends). Thus the value of equity can be represented either as the present value of a growing dividend stream or as the present value of a constant earnings stream. The fact that $p$ does not depend on $\delta$ reflects the Miller-Modigliani proposition (Merton Miller and Franco Modigliani [70]) that (in this simple environment) for given current earnings the future dividend payout rate does not affect current equity value.
3.2 A Stochastic Version

For the present purpose, the main shortcoming of the Gordon model is that it is deterministic. As such, it provides no insight into stock price fluctuations. To remedy this deficiency we replace the assumption that the earnings growth rate is constant with the assumption that the earnings growth rate is independently and identically distributed over time or, put differently, that earnings follow a geometric random walk.\(^2\) We refer to the model that results from this generalization as the stochastic Gordon model.

The first four autocorrelations of the earnings growth rate are 0.050, −0.145, −0.340 and −0.078 (annual data, 1958 to 2000), so one would not want to defend too strongly the assumption that earnings growth is white noise. The predominantly negative autocorrelations imply that earnings have a mean-reverting component. To the extent that investors take this mean-reversion into account in valuing stocks, the model to be presented gives an upward-biased account of stock price volatility.

Figure 4 shows the interest rate less the growth rate of earnings; the series looks like an IID process, as assumed in the stochastic Gordon model.\(^3\) Figure 5 shows the dividends-earnings ratio. Obviously the assumption adopted in the stochastic

\(^2\)The geometric random walk was found more that 40 years ago to give an accurate description of corporate earnings in the UK (A. C. Rayner and I. M. D. Little [77], W. B. Reddaway [78]).

\(^3\)The attractive feature of this variable is that we do not have to worry about inflation adjustment, since the inflation correction involved in figuring real earnings growth offsets that involved in calculating the real interest rate.
The more information about the future that investors have, the more volatile will be the price-earnings ratio. This will be demonstrated below, but the idea is intuitive: if investors have no information about future earnings, they will price stock at a constant multiple of current earnings, so the price-earnings ratio will have zero variance. To the extent that investors have information about future earnings realizations, they will
price stocks at varying multiples of current earnings. An upper bound on the variance of the price-dividend ratio will be generated in the case of complete knowledge of future earnings. Stephen F. LeRoy and William R. Parke [60] showed\(^4\) that this variance is given by the right-hand side of the following inequality, so the variance bound is
\[
V(p/e) \leq \frac{\beta^2 \delta^2 \sigma^2}{(1 - \beta^2(\mu^2 + \sigma^2))(1 - \beta \mu)^2}.
\] (2)

Here \(\beta\) is the discount rate, and \(\mu\) and \(\sigma^2\) are the mean and variance, respectively, of the earnings growth rate. Inserting estimated parameter values and converting to standard deviations, the upper bound on price volatility is 7.274, whereas the standard deviation of the actual price-earnings ratio is 8.945. Thus the point estimates indicate some excess volatility.

3.2.2 Volatility of Returns

Just as increasing information increases the volatility of the price-earnings ratio, it decreases the volatility of the rate of return: in the extreme case in which investors know the entire path of future earnings, the volatility of the rate of return will be zero. Accordingly, the case in which investors have no information beyond current earnings will generate an upper bound on the variance of the rate of return.

If investors have no information about future earnings growth rates, the variance of the rate of return under the geometric random walk model is given by the right-hand side of the following inequality. Therefore we have
\[
V(r) \leq \frac{\sigma^2}{\beta^2 \mu^2}.
\] (3)

Inequality (3) is the basis for Kenneth D. West’s [92] test, adapted to the geometric random walk case. It is essentially the same as LeRoy and Richard D. Porter’s [61] lower bound on price volatility. Based on estimated parameter values and again converting to standard deviations, the upper bound is 0.179, whereas the standard deviation of actual returns is 0.174. Thus return volatility is almost exactly what it would be under the stochastic Gordon model if investors had no information about future earnings growth rates.

3.2.3 Price-Earnings and Return Volatility

The foregoing discussion suggests that there is a tradeoff between price volatility and return volatility: high price volatility is associated with low return volatility and

\(^4\)This and several other expressions in the text differ in detail from those in LeRoy-Parke because those authors were working with price-dividend ratios rather than price-earnings ratios.
vice-versa. In the setting of the stochastic Gordon model there exists a closed-form expression for the locus of pairs of price variance and return variance as a function of parameters:

\[ V(p^*/e) = V(p/e) + \frac{\beta^2[V(p/e) + (\beta\delta\mu/(1 - \beta\mu))^2]}{(1 - \beta^2(\mu^2 + \sigma^2))}V(r) \]  \hspace{1cm} (4)

Here \( V(p^*/e) \) equals the upper bound of the variance of the price-earnings ratio (the right-hand side of (2)). Incidentally, equation (4) does not require that earnings follow a geometric random walk; the role of the random walk assumption is to provide an expression—the right-hand side of (2)—for \( V(p^*/e) \).

Inserting values for \( V(p/e) \) in (4) and solving for \( V(r) \) allows us to compute the locus of pairs of price volatility and return volatility that are consistent with the Gordon model, given the mean and variance of the earnings growth rate, the discount rate and the dividend payout rate. Equality (4) is essentially the same as the relation between price volatility and payoff volatility derived by LeRoy and Porter [61], adapted to the geometric random walk case.

Figure 6 shows the locus of standard deviations of the price-earnings ratios and rates of return that are consistent with estimated parameter values (standard deviations are easier to interpret than variances). The diagram displays a dot representing the estimated standard deviations from US stock data. As the diagram makes clear, the volatility measures are higher than the stochastic Gordon model implies.\(^5\)

\(^5\)Some versions of the variance-bounds tests have been subjected to criticism on econometric
3.3 The 1990s Price Runup

The variance bounds tests just presented demonstrate that, overall, price and return volatility exceed that implied by the stochastic Gordon model. Inspection of the diagrams makes clear that the principal source of both price and return volatility is the stock price runup in the 1990s and the subsequent price collapse. Can this specific episode be interpreted within the framework of the stochastic Gordon model? Under the Gordon model, price-earnings ratios will be unusually high when investors have information that leads them to predict extranormal future earnings increases. As the data indicate, earnings were in fact increasing during the late 1990s, and it is reasonable to suppose that investors were extrapolating these increases into the future. However, the increases in price-earnings ratios exceeded by orders of magnitude anything justifiable under the Gordon model, barring wildly optimistic earnings forecasts.

In a series of papers Robert E. Hall examined stock prices in the late 1990s from the vantage of a one-good production model (Hall [43], [44], [45], [46]), an environment that differs in some respects from the Gordon model. In the simplest such setting (that is, in the absence of adjustment costs), the value of tangible corporate capital increases by the amount of investment net of depreciation. As Hall showed, investment was insufficient to explain the equity price runup.

Intangible corporate capital, however, is less easily dismissed. Hall proposed that the advent of the internet led to a creation of new intangible capital on the order of several trillion dollars, and this new capital constituted the basis for the stock price runup. How this capital came into existence is unclear, and the very modest increase in aggregate corporate earnings in the 1990s suggests either that a large component of the return on this new capital somehow went unmeasured or that the return per unit of capital suffered a sharp drop. Finally, the stock price declines of the last several years imply that this new capital disappeared even more rapidly than it had materialized.

In the absence of any evidence in favor of this hypothesis other than the fact that a huge change in stock prices occurred unaccompanied by anything like a commensurate change in measured corporate earnings, it is difficult to believe that variations in intangible corporate capital played a major role.

3.4 Evaluation

The stochastic Gordon model is simple and plausible, but it does not provide a good explanation for observed stock price behavior: it predicts less volatile prices grounds (see, for example, Christian Gilles and LeRoy [36] and the papers cited there). At one point it appeared, to me at least, that there was reason to expect that the excess volatility of security prices could be explained as a consequence of econometric problems (LeRoy [58]). However, it subsequently became clear that variance bounds tests not subject to major econometric problems, such as those just outlined, still show excess price volatility.
and returns than we see, and it does not give useful guidance in thinking about the episode of the 1990s. From a theoretical point of view, the stochastic Gordon model has two major shortcomings: (1) it assumes a constant discount factor, and (2) it excludes bubbles. In the following section we will consider the generalization to stochastic discount factors under the rubric of consumption-based asset pricing.

4 Consumption-Based Asset Pricing

In stationary exchange economies with risk-neutral agents, asset prices will equal expected payoffs discounted at the risk-free interest rate, as assumed in the stochastic Gordon model. However, when agents are risk averse, expected returns will be different for different securities depending on their risk. The theory of consumption-based asset pricing, introduced in the economics literature by Robert Lucas [63] and in the finance literature by Douglas Breeden [13], shows how asset valuation is altered by risk aversion.

Expositions of consumption-based asset pricing can be found in a variety of sources (for example Peter L. Bossaerts [12], John C. Cochrane [22], LeRoy and Jan Werner [62], or John Y. Campbell, Andrew W. Lo and A. Craig MacKinlay [18]; the last of these gives a very clear discussion of the empirical tests of consumption-based asset pricing).

If agents' utilities are strictly concave functions of their consumption levels, equilibrium expected returns will depend on how asset payoffs covary with aggregate consumption. Risk-free returns and risk premia, both of which are assumed constant in the stochastic Gordon model, will be stochastic, and the latter will be different for different assets. Thus there is no reason to regard the variance bounds tests as applying in a world of risk-averse agents.

In fact, there are specific reasons to expect that stock price volatility will exceed that implied by risk neutrality when agents are risk averse. When agents are risk averse they will be motivated to use portfolio transactions to smooth consumption. In a representative agent exchange setting, they cannot do so in the aggregate. That being the case, asset prices must be such as to induce the representative agent to consume his endowment. As a result, equilibrium asset prices will covary positively with consumption, and the more risk averse agents are, the higher the covariance must be. Thus risk aversion creates a presumption in favor of increased price volatility relative to risk neutrality. C. J. LaCivita and LeRoy [55] formalized this argument; a related development is Sanford J. Grossman and Robert J. Shiller [42].

As is well known, the problem with consumption-based asset pricing theory is that it appears to be a failure empirically. The implication that expected asset returns depend on covariances with consumption growth (or, in the market CAPM version, with the return on the market portfolio) appears not to be satisfied in the data. Major papers demonstrating the failure of market CAPM and consumption CAPM are Lars P. Hansen and Kenneth J. Singleton [48], Rajnish Mehra and Edward C. Prescott [69],
Hansen and Ravi Jagannathan [47] and Eugene F. Fama and Kenneth R. French [29]. Given the failure of consumption-based asset pricing theory to provide an empirically successful account of variations in expected returns over time and across securities, we will proceed henceforth as if expected returns are constant.6

It is possible that the problem lies not with consumption-based asset pricing theory in general, but with its implementation using expected utility theory. A variety of difficulties with expected utility theory have been documented. The most recent of such arguments is that expected utility theory appears to give implausible predictions when applied to large and small gambles (see Matthew Rabin [73] and Rabin and Richard H. Thaler [74]).

Many analysts (Cochrane [21], for example) conclude from the failure of consumption-based asset pricing theory that the quest for a successful implementation of the stochastic discount factor must continue, since we know that investors are risk averse. A variety of ever more elaborate specifications are being tried, ranging from non-expected utility to transactions costs and various forms of market incompleteness. So far none of these modifications appears to have succeeded, but the jury is still out. For the present, we can conclude that the fact that expected asset returns do not covary with aggregate consumption means that we lose little in terms of explanatory power by suppressing the dependence of discount factors on consumption. This amounts to accepting the assumption of the stochastic Gordon model that discount factors are constant.

5 Bubbles: A Taxonomy

In the financial media the question “Do you think internet stocks (real estate, Japanese stocks) are or were a bubble?” appears to mean that the questioner wants to know whether prices will collapse any time soon. Along these lines, a bubble means simply a big price rise that is shortly followed by an equally big price drop. Although the argument is seldom spelled out explicitly, the presumption is that bubbles should be avoided as investments because the probability of a collapse exceeds that of a further rise (allowing for the effects of risk-aversion and discounting). Under rational asset pricing, including rational expectations, such biased expectations cannot occur: absence of arbitrage implies that the expected (risk-adjusted and discounted) gain on any security or portfolio is zero. Thus in this usage bubbles are synonymous with

---

6 The alternative is to specify directly a time series model for returns that involves intertemporal dependence, but not attempt to relate the innovations in expected returns to aggregate consumption. In this setting the restriction to constant returns is clearly rejected empirically at customary significance levels (see, for example, Campbell, Lo and MacKinlay [18]). The problem with this specification is that it involves assuming not only that (conditional) expected returns are time-varying, but that investors make allowance for this dependence in pricing securities. This may not be plausible. Similar reservations about indiscriminately invoking rational expectations are recorded in Section 6 below.
irrationality.

In professional discussions “bubble” has several meanings that must be distinguished. At the cost of some oversimplification, four distinct meanings may be identified. These are summarized in the four subsections that follow. First and most obviously, the identification of bubbles with irrationality can be carried over. In the older academic literature this identification is taken for granted. The same is true in some contemporary discussions. For example, Garber [33] expressed the opinion that

Bubble is one of the most beautiful concepts in economics and finance in that it is a fuzzy word filled with import but lacking a solid operational definition. Thus, one can make whatever one wants of it. The definition of bubble most often used in economic research is that part of asset price movement that is unexplainable based on what we call fundamentals.

Garber questioned the presumption implicit in these accounts that fundamentals cannot explain even such episodes as the tulip bulb speculation in Holland in the seventeenth century (Garber’s book is reviewed in LeRoy [59]). Similarly, Hall [45] wrote

I reject market irrationality in favor of the hypothesis that the financial claims on firms command values approximately equal to the discounted future returns.

Failure of prices to equal discounted returns is taken here to be equivalent to irrationality, so there is no allowance for the possibility that values could exceed discounted future returns when agents are fully rational.7

Bubbles as manifestations of irrationality are considered in the first subsection below. The remaining subsections of this section outline three distinct rational-agent settings in which bubbles may be defined and analyzed. In the second and third subsections we abstract from asymmetric information. In the second subsection a setting of simultaneous markets is adopted, while in the third subsection sequential markets are specified. In the fourth subsection models are reviewed that allow analysis of bubbles under asymmetric information.

In this section we make the preliminary observation that rational bubbles can occur only in infinite-date models. Extended consideration of conditions under which bubbles can and cannot occur in these classes of models is deferred to the following section.

---

7 Later in his paper Hall acknowledged the possibility of rational bubbles, but dismissed them on the grounds that existence of such bubbles would violate “a fundamental efficiency condition”.

This argument is examined below.
5.1 Irrational Bubbles

The meaning of “rationality” and “irrationality” in economic discussion appears to have changed in recent years. In mainstream nonfinancial economic theory the received view has been that rationality is not a substantive hypothesis about the world, but rather a conceptual tool used in formulating economic models. In fact, the idea of rationality provides the basis for a working definition of an economic model as a model describing human (or, for that matter, animal) behavior that assumes consistent choice. Along these lines there is no room for irrational behavior: apparently irrational behavior is actually evidence of omitted costs or the like. As this argument makes clear, taking rationality as an analytical tool rather than a substantive hypothesis gives rationality a tautologous character. As part of a maintained hypothesis, no conceivable evidence can contradict it.

A generation ago financial economists fully accepted this view of rationality, if the term “market efficiency” is substituted for “rationality”. In his classic survey of efficient capital markets Fama [28] emphasized that market efficiency by itself has no observable implications. It can be tested only in conjunction with a particular market model.8

Contemporary exponents of behavioral finance are more sympathetic to the idea of irrationality than mainstream financial economists were a generation ago. However, in making their point that irrationality plays a greater role in financial markets than hitherto believed, exponents of behavioral finance have altered the definition of rationality. Irrationality is identified not with nontransitivity, but with the existence of agents who trade for reasons that are not modeled (“noise traders”, introduced into the finance literature by Albert S. Kyle [54] and Fischer Black [9]).

Models in which some of the economic behavior being modeled is taken as given, as distinguished from models in which all agents are represented as optimizing subject to constraint, have been known in the received economics literature as partial equilibrium models. There was no suggestion that partial equilibrium has anything to do with irrationality. It is difficult to see the advantage in broadening—and blurring—the definition of irrationality as the advocates of behavioral finance have done, but there is little doubt that this has happened.

In the behavioral finance literature the standard vehicle for analyzing the effects of irrationality (under the expanded definition) is a model in which there exist both noise traders and rational traders (for example, Andrei Shleifer [85]). Oddly, this setting coincides exactly with that envisioned in the early finance literature as embodying efficient markets. It was emphasized there that market efficiency does not require

---

8To be sure, in the early empirical literature on market efficiency financial economists, Fama included, were not always clear about exactly what market model was embedded in the hypothesis being tested. Accordingly, they often described empirical results as favoring or not favoring market efficiency directly, a usage that conflicts with Fama’s dictum. Also, it should be noted that Fama’s formulation of market efficiency was tautologous in a sense different from that discussed above, as I pointed out in LeRoy [57].
exclusion of noise traders as long as there are enough rational agents to dominate asset pricing (formally, this means that in equilibrium the portfolios of rational agents are interior). Proponents of market efficiency emphasized that in the setting just described the major propositions of neoclassical finance theory, such as the risk-return tradeoff, will survive the presence of noise traders, implying nonexistence of profitable trading rules.

As noted, Shleifer and other proponents of behavioral finance adopted exactly the same analytical setup, but interpreted it as reflecting inefficient rather than efficient capital markets. Rather than identifying market efficiency with the risk-return tradeoff (equivalently, the nonexistence of profitable trading rules), they emphasized a different aspect of equilibria in models populated by both rational and irrational agents: that in general the existence of noise traders will affect asset prices in equilibrium even if the rational agents are at interior optima. This is so because if the rational agents are risk averse, they will rationally reject trades that exploit minor mispricing because of the added risk, implying that they will generally not completely eliminate the effects of noise traders on security prices (John Maynard Keynes [52], p. 157, J. Bradford De Long et al. [25]). In the early literature on market efficiency it was presumed (although generally not stated) that the rational agents would completely arbitrage away the effects of noise trades on security prices. This conclusion, we now know, follows only if agents are risk neutral, a specification often made implicitly in early discussions of market efficiency (LeRoy [56], [57]). This is so because if all agents are risk neutral, then rational traders will bid all asset prices to levels that equate expected returns, so the rational traders completely eliminate the effect of noise traders on asset prices.

Let us specialize this discussion to bubbles. In the literature associating bubbles with irrationality, bubbles are identified with the herd behavior of noise traders, who bid prices above levels that can be justified by any rational calculation. Rational agents perceive the mispricing and bet against it but, assuming that they are risk averse, their trades are not sufficient to eliminate the mispricing.

In the case of the 1990s US stock price runup, it was easy for any trader to bet against the bubble by, for example, selling NASDAQ futures or buying puts. Few perceived doing so as an attractive trading opportunity for the obvious reason that the mispricing might increase, implying losses. Investors who made this bet too early did in fact post losses, often substantial. In particular, it appears that institutional traders, who presumably correspond to the rational agents of the models, typically bet with the bubble rather than against it (see, for example, John M. Griffin et al. [41]).

The fact that virtually all of the traders who believed that stocks were overvalued still declined to bet against the bubble might just mean that there exist very few rational agents. A more likely possibility, we argue, is that betting against the bubble is not profitable, so the fact that few traders did so does not contradict the proposition that most investors trade rationally. This points toward rational bubbles,
analyzed below.

Two qualifications are necessary. First, papers like that of Dilip Abreu and Markus K. Brunnermeier [2] presented models in which rational traders bet with the bubble rather than against it. To the extent that these models are valid, failure to identify rational traders who bet against the model does not count against the hypothesis of irrational bubbles. Second, in asset price runups prior to the 1990s, index futures and options did not exist, implying that it was impossible to take a short position against the market as a whole. To be sure, investors could take short positions in individual issues that they believed to be overpriced. However, in doing so they ran the risk of being subjected to a short squeeze, which was a regular feature of early stock price manipulations (Chancellor [19]). Therefore the foregoing argument against the irrationality of the 1990s bubble does not necessarily apply to earlier episodes. Also, it does not apply to chain letters and pyramid schemes, which cannot be sold short.

In sum, if bubbles reflect the actions of noise traders, all rational traders will trade against the bubble (subject to the qualification just stated). In so doing they will experience better outcomes (in expected utility terms) than the noise traders. In contrast, under rational bubbles returns on assets with bubbles have the same probabilistic characteristics as those on assets without bubbles, so there is no presumption that rationality implies betting against the bubble.

A number of recent papers model bubbles in a setting that does not presume full rationality and rational expectations, but also avoids the noise trader construction (that is, these papers avoid including in the model traders whose behavior is taken as given rather than explained in terms of optimization). Instead, the traders are assumed to optimize, but also to have expectations that are biased in some way. For example, De Long et al. [26] analyzed a setting in which traders are systematically optimistic, whereas Jose Scheinkman and Wei Xiong [81] assumed a setting in which agents place more weight on their own information than its accuracy justifies.

5.2 Rational Bubbles: Simultaneous Markets

In the preceding subsection bubbles were implicitly defined as the difference between equilibrium asset prices and those that would prevail in the absence of noise traders. That definition is obviously unsuitable if there are no noise traders. Rather, the bubble is defined as the difference between the equilibrium price of an asset and the present value of its finite-date dividends. Specifically, suppose for convenience that the setting is deterministic and stationary (relaxing either of these restrictions complicates the notation but adds nothing of substance). Then if some security or portfolio with date-$t$ dividend $d_t$ is held for $n$ periods and sold, its purchase price $p_t$
and sale price \( p_{t+n} \) must obey

\[
p_t = \sum_{i=1}^{n} (1 + r)^{-i} d_{t+i} + (1 + r)^{-n} p_{t+n}
\]  

(5)

in the absence of arbitrage. Here \( r \) is the one-period net interest rate, assumed constant, and \( d \) is dividends. Allowing \( n \) to go to infinity, (5) becomes

\[
p_t = \sum_{i=1}^{\infty} (1 + r)^{-i} d_{t+i} + \lim_{n \to \infty} (1 + r)^{-n} p_{t+n},
\]  

(6)

assuming that the limits exist. The first term is \( f_t \), the fundamental value of the security—its value based on finite-date payoffs—and the second is \( b_t \), its bubble. We have

\[
p_t = f_t + b_t.
\]  

(7)

By substituting (7) and the definitions of \( f_t \) and \( b_t \) in the no-arbitrage condition

\[
p_t = (1 + r)^{-1}(p_{t+1} + d_{t+1}),
\]  

(8)

there results

\[
b_t = (1 + r)^{-1}b_{t+1},
\]  

(9)

so bubbles, if they exist at all, grow at the interest rate. In a setting of uncertainty the analogous condition is

\[
b_t = (1 + r_t)^{-1}E_t(b_{t+1}),
\]  

(10)

where \( E_t \) denotes conditional expectation with respect to a probability measure that makes allowance for the effects of risk aversion. Thus under uncertainty the prospect of high returns on bubbles if they do not burst just offsets the possibility of their bursting (see Olivier Blanchard and Mark Watson [11] or Blanchard and Stanley Fischer [10], Chapter 5 for further discussion).

When can we exclude the possibility that (nonzero) bubbles exist? Certainly if the number of dates is finite there cannot exist a bubble at the last date, and a backward induction argument (based on (9)) implies that the bubble must equal zero at every earlier date as well. If the number of dates is infinite, this argument cannot be used because there is no last date.

Jean Tirole [89] asserted that even if the future is infinite, bubbles can always be excluded as long as the number of agents is finite. This assertion requires qualification, and the nature of the qualification depends on how infinite-date settings are modeled.

One can assume either simultaneous or sequential markets (the introductory material on bubbles presented up to this point applies in both settings). Under simultaneous markets agents are viewed as trading once and for all on an infinite set of
markets that meet before time begins (and on which contingent claims are traded in the case of uncertainly). Discussion of sequential markets is deferred to the following subsection. The two frameworks are equivalent in finite settings (Kenneth J. Arrow [6]), but not in the infinite settings that are of interest here. In the following section, which discusses conditions under which bubbles can be shown not to occur, it is shown that the analysis of bubbles is quite different in the two cases.

Under simultaneous markets, one proceeds by applying received static general equilibrium theory to a multidate setting: equilibrium consists of a linear price function that assigns value to commodity bundles, just as in nonfinancial applications, and a set of optimal consumption bundles that clear markets. The fact that time is infinite means that consumption sets are infinite-dimensional. Infinite-dimensional consumption sets have some inconvenient implications (proving existence of equilibrium becomes more complicated, for example; see A. Araujo [5]), but these problems are not relevant here.

Example 1. To see how bubbles can occur under linear pricing functions, consider the pricing function

\[ p(y) = \sum_{t=1}^{\infty} (1 + r)^{-i} y_t + \inf_{n} ((1 + r)^{-n} y_n), \]  

(11)

where \( y = \{y_t\} \) is any payoff, and define date-\( t \) prices by

\[ p_t = (1 + r)^t p(y\{t\}), \]  

(12)

where

\[ y\{t\} = \begin{cases} 0 & \tau \leq t \\ y_{\tau} & \tau > t. \end{cases} \]  

(13)

From (12) and (13), the price of a security at \( t \) equals the value of its future payoffs expressed in units of consumption at date \( t \). Linearity of \( p \) is easily verified, and from the definitions of fundamental and bubble it is clear that the first term on the right-hand side of (11) is the fundamental value associated with dividend \( y \) and the second term is its bubble. In the Section 6 we will be considering when price systems like (11) can occur in equilibrium.

5.3 Rational Bubbles: Sequential Markets

We have seen that an attractive feature of specifying simultaneous markets in the analysis of bubbles is that the required departures from classical general equilibrium theory are minor, consisting of little more than respecifying commodity and price
spaces to be infinite-dimensional. However, simultaneous-markets models have disadvantages: the assumption that all trading occurs simultaneously strikes many economists as implausible, particularly when applied in an infinite-date setting.

An alternative is to specify sequential markets. General references on sequential markets are Michael Magill and Martine Quinzii, [65], [66], [67] and John Geanakoplos [34]. Under sequential markets agents trade at each date rather than simultaneously, implying that they face separate budget constraints at each date and event. These budget constraints specify that each agent’s consumption does not exceed his endowment minus the value of net acquisitions of securities. Thus in place of the integrated budget constraint that applies under simultaneous markets, we have the sequence of budget constraints

\[ c_t \leq x_t - \theta_t p_t + \theta_{t-1} (p_t + d_t). \]  

(14)

Here \( x_t \) is the agent’s endowment and \( \theta_t \) is a vector of which the \( j \)-th element is the holding of the \( j \)-th security. Similarly, \( p_t \) and \( d_t \) are vectors of security prices and dividends.

Sequential markets involve a serious difficulty. Agents’ decisions at each date depend not only on the security prices they face at that date, but also on the probability distribution of security prices in the future. Where, under sequential markets, do these probability distributions for future prices come from? Following Roy Radner [75], most analysts simply assume that agents know prices at future events. This assumption is usually defended by a loose appeal to rational expectations, the plausibility of which depends on how much recursive structure the economy has. Under simultaneous markets there is no corresponding problem because the fiction of the Walrasian auctioneer, who provides information about future as well as present prices, is maintained. However, the appeal of sequential markets is exactly that it dispenses with the auctioneer in favor of a specification that is thought to be more realistic.

The point of Arrow’s classic paper [6] is that in finite settings successive retring of a few securities can duplicate the (Pareto-optimal, if markets are complete) allocation achieved in simultaneous markets in which agents trade many contingent claims. This result, however, depends critically on the assumption that agents know future prices, a fact which greatly diminishes its appeal. Under simultaneous markets many contingent claims prices are needed to achieve efficient allocations precisely because efficient resource allocation requires measures of scarcity along many dimensions. Replacing many contingent claims with a few retrained securities results in the

\[ \text{It could be argued that the term “classical general equilibrium theory” should include the infinite-dimensional case, which was studied as early as the 1950s (Gerard Debreu [23]), the same time as standard general equilibrium theory as reflected in Debreu [24] was being developed.} \]

\[ \text{Arrow’s result leads one to ask under what conditions equilibrium allocations under simultaneous markets coincide with those implied by sequential markets. Several papers that discuss this issue are Randall Wright [93], M. Kandori [51], Darrell Duffie and Chi-fu Huang [27] and Kevin Huang and Werner [49], [50].} \]
same allocation only if the information that would be communicated by the many contingent claims prices is somehow otherwise available. Arrow gave no account of how this happens. Accordingly, his claim that a few securities can do the same work as many contingent claims is questionable. This criticism of Arrow’s result is due to Keizo Nagatani [71].

These considerations raise questions about whether the specification of sequential markets is in fact more attractive than that of simultaneous markets: both specifications appear to rest on assumptions that are hard to defend. The fact that the two specifications are equivalent in finite settings means that a preference for one over the other in infinite settings must reflect some intuitive idea of how people behave when the future has no finite endpoint. One would like to see the widely-held idea that sequential markets are more plausible than simultaneous markets defended explicitly.

The form of the budget constraints in sequential models implies that in the absence of trading restrictions agents can operate Ponzi schemes—borrow money, spend the proceeds and roll over the indebtedness forever. (Under simultaneous markets, on the other hand, there is no analogue to Ponzi schemes because the budget constraint directly restricts the value of each agent’s consumption to at most equal the value of his or her wealth.) As arbitrages, Ponzi schemes are inconsistent with existence of optima. Models analyzing bubbles under sequential markets therefore necessarily impose trading restrictions. The trading restrictions required to rule out Ponzi schemes may also prevent agents from implementing the arbitrages that they would otherwise use to exploit bubbles. If so, bubbles can exist in equilibrium.

Example 2: Suppose that portfolio transactions are subject to the wealth constraint, which restricts the value of an agent’s short sales to the present value of his future endowment. Assume also that there exists a single security with payoff proportional to the aggregate endowment, and that all agents have zero endowments of this security. If the claim on the aggregate endowment has a bubble, traders want to sell this security and roll over the short position forever. They would use the proceeds either to consume immediately or buy future consumption on spot markets. Since the bubble increases in value at the rate of interest, at some future date the value of the short position will necessarily eventually exceed the value of the remaining endowment. Agents would have to close out the short position, meaning that they could not complete the purported arbitrage. The infeasibility of this transaction means that nothing prevents the bubble from occurring in equilibrium. This bubble is inconsequential in that its existence does not affect equilibrium consumption allocations, but it is a bubble nonetheless.

Example 3: As is well known, in overlapping generations models intrinsically useless money can have positive value in equilibrium (Neil Wallace
If money is intrinsically useless its fundamental value is zero, so valued money is an example of a bubble (Tirole [90]). If agents could sell money short and roll over the short position forever, they would do so, thereby arbitraging away the bubble. Monetary authorities therefore necessarily impose limitations on issue of money, prohibitions on counterfeiting and the like to rule out the transactions that would otherwise exploit the bubble. Also, in overlapping generations models it is assumed, reasonably, that agents can trade only during their finite lifetimes, implying that they cannot maintain short positions in money into the indefinite future.

5.4 Rational Bubbles: Asymmetric Information

The papers summarized up to now analyzed bubbles in settings that abstract away from asymmetric information. Recognizing that individuals have different information, and therefore different probability distributions for future security prices and payoffs, means that the division of asset value into fundamental and bubble is no longer unambiguous: in general agents will disagree about whether a particular security has a bubble or not.

Several recent authors have produced models in which asymmetric information plays a central role, but in which security prices clearly have bubbles (under reasonable definitions of bubbles). Markus K. Brunnermeier [15] is an excellent introduction to bubbles, and asset pricing generally, under asymmetric information. Franklin Allen and Gary Gorton [3], for example, modeled trading in a security that is known to have zero value at a terminal date, and therefore has zero fundamental value. There are three traders, the first of which has the security as his endowment. The first trader sells the security to the second trader, who in turn sells it to the third trader for a profit. The third trader loses his or her investment when the terminal date arrives and the security is revealed to have zero value.

Two aspects of the model specification make this game of musical chairs work as an equilibrium: (1) the assumption that the second and third traders do not know their identities, and (2) the assumption that traders have convex payoff functions. Because neither the second nor the third trader knows which trader he is, each is willing to buy the security in the hope that he is the second trader and is buying from the first trader, in which case he will sell to the third trader at a profit. It is assumed also that these traders, like real-world hedge fund managers, retain a portion of any profit, but losses are born entirely by investors. Without this provision the traders would never pay a positive price for the security. Since the security pays zero dividends and will ultimately have zero value, its fundamental value is clearly zero. Therefore the equilibrium price consists entirely of a bubble.

We are classifying Allen-Gorton’s model as a rational bubble, although one might question this characterization because of the dependence of the results on the heads-
I-win-tails-you-lose compensation of the traders. This leaves open the question of why the investors, who lose money on average, participate. However that may be, an attractive feature of Allen-Gorton’s model is that it illustrates, although in a highly stylized setting, the possible role of asymmetric information and agency problems in creating bubbles.

Allen, Stephen Morris and Andrew Postlewaite [4] contributed a model in which each of three agents has a different information partition, and each agent extracts information optimally from equilibrium prices. They produced an example of a rational expectations equilibrium in which there exists a state in which each agent knows that the value of the security exceeds its payoff. However, that fact is not common knowledge: because of the asymmetric information, each agent does not know that the other agents know that the asset is overpriced. The authors suggest interpreting the fact that each agent knows that the asset is overpriced as constituting a bubble. Doing so raises questions: should a security be defined to have zero fundamental value if it is common knowledge that its payoff is zero (which is not true in Allen-Morris-Postlewaite’s example) or should it be required only that each agent knows that its payoff is zero (which is true in the example)? The authors are well aware of such difficulties, and provide very interesting discussion.

5.5 Evaluation

We have outlined four settings in which bubbles may be analyzed: (1) agents act irrationally, (2) agents have identical information and trade in simultaneous markets, (3) agents have identical information and trade in sequential markets, and (4) agents have asymmetric information. Settings (2) and (3) will be retained for further discussion in the next section. As regards irrational bubbles, we have already suggested that the identification of irrationality with noise traders is misdirected: attributing bubbles to unexplained behavior does not qualify as explanation. Introducing unexplained biases into agents’ expectations-formation process raises similar issues: why do such biases exist and, especially, why do they persist in equilibrium?

As regards asymmetric information there exist no such methodological problems. These models are not considered further because they simply do not lend themselves to applied work. As difficult as such questions as “Can bubbles explain excess volatility of stock prices?” are under rational expectations with symmetric information, they are virtually impossible under asymmetric information.

6 Conditions that Rule Out Bubbles

In this section we analyze rational bubbles in the two preferred settings: simultaneous markets and sequential markets. As emphasized by Kevin Huang and Werner [49] and noted above, the analysis of bubbles is different in the two cases.
6.1 Simultaneous Markets

We observed in Subsection 5.2 that bubbles cannot occur in finite-time settings. Here we show that bubbles can sometimes be excluded even in infinite-time settings under simultaneous markets. The following example is a general-equilibrium version of the Gordon model:

Example 4. Consider a deterministic exchange economy in which the representative agent has utility function

\[ U(c) = \sum_{t=1}^{\infty} \beta^t c_t. \]  

(15)

The representative agent’s endowment \( x_t \) grows at a constant rate \( g \):

\[ x_t = x_0 (1 + g)^t, \]  

(16)

where \( \beta (1 + g) < 1 \) to ensure that consuming the aggregate endowment generates finite utility. With \( g = 0 \) we have the Lucas [63] tree economy, specialized to a deterministic setting.

The representative agent’s budget constraint is

\[ \sum_{t=1}^{\infty} (1 + r)^{-t} (c_t - x_t) \leq 0, \]  

(17)

anticipating that in equilibrium the interest rate will be constant. If there is no bubble, the equilibrium price system \( p \) is given by

\[ p(y) = \sum_{t=1}^{\infty} (1 + r)^{-t} y_t \]  

(18)

where the equilibrium interest rate \( r \) is related to preferences according to

\[ \beta = (1 + r)^{-1}. \]  

(19)

If (19) failed the representative agent would reject consuming his endowment in favor of borrowing or lending, depending on the direction of inequality. This cannot occur in equilibrium.

Measured in units of date-\( t \) consumption, the date-\( t \) value of the aggregate endowment, as the capitalized value of a growing perpetuity, satisfies the Gordon equation

\[ p_t = \frac{1 + g}{r - g} x_t, \]  

(20)
assuming that there is no bubble. If there were a bubble on the aggregate endowment with date-0 value \( b_0 \), (20) generalizes to

\[
p_t = \frac{1 + g}{r - g} x_t + b_0 (1 + r)^t.
\]  (21)

For this economy we can state definitively that the aggregate endowment cannot have a bubble: \( b_0 = 0 \). If on the contrary \( b_0 > 0 \), the representative agent would want to sell the aggregate endowment at date 0 and consume the proceeds of the sale immediately. Doing so would increase period utility at date 0 by \((1 + g)x_0 / (r - g) + b_0\) and would decrease discounted future utility by \((1 + g)x_0 / (r - g)\), for a net gain of \( b_0 \).

Formally, what is happening here is that optimal portfolio choice implies not only the usual first-order conditions that are necessary in finite-dimensional optimizations, but also a necessary transversality condition. A zero net trade (which must be optimal in any representative-agent environment) violates the transversality condition in the presence of a nonzero bubble. This argument eliminating bubbles appears to be due to Maurice Obstfeld and Kenneth Rogoff [72].

It is easy to draw from this argument the incorrect conclusion that bubbles cannot occur in any representative agent setting. To see the error, consider the following modification of the example:

**Example 5:** Change the representative agent’s utility function to

\[
U(c) = \sum_{t=1}^{\infty} \beta^t c_t + \inf_n ((1 + g)^{-n} c_n).
\]  (22)

The rightmost term in (22) specifies that agents derive utility from their minimum consumption level, discounted at rate \( g \), as well as the period-by-period consumptions. Note here that \( g \) is exactly the rate at which the aggregate endowment increases. This utility function, while unusual, is increasing, concave and incorporates discounting, so it cannot be dismissed out of hand.\(^{12}\)

\(^{12}\)Graciela Chichilnisky [20] suggested that utility functions like (22) are more suitable than standard utility functions like (15) for evaluating environmental projects. This is so because the costs and benefits of such projects extend over many generations; discounted utility with discount factors on the order of two or three per cent per year (the average real return on financial assets) assign essentially zero weight to the welfare of generations after the first few. Environmentalists, following Frank Ramsey [76], find this unacceptable. Chichilnisky suggested using utility functions like (22) to represent “sustainable preferences.” Thus if agents have sustainable preferences, asset prices will have bubbles.
The equilibrium price system associated with the utility function (22) is
\[ p(y) = \sum_{t=1}^{\infty} \beta^t y_t + \inf_n ((1 + g)^{-n} y_n). \] (23)

In particular, the aggregate endowment \( x \) is valued according to
\[ p(x) = \frac{x_0 (1 + g)}{r - g} + x_0, \] (24)
as is verified by substituting (16) in (23). Here \( x_0 (1 + g)/(r - g) \) is the fundamental value of the aggregate endowment, while \( x_0 \) is its bubble. With the utility function (22) and price system (23), the representative agent does not gain by selling the aggregate endowment and consuming the proceeds immediately. This is so because the loss of utility from future consumption is \((1 + g)x_0/(r - g) + b_0\), not \((1 + g)x_0/(r - g)\) as in the preceding case.

This example is obviously contrived in that the discount factor \( g \) in the rightmost term of (23) just equals the growth rate of the aggregate endowment. This is a knife-edge case: if the aggregate endowment increases at a rate greater than the rate at which agents discount the future, it will have infinite value, so there can exist no equilibrium. If, on the other hand, the growth rate of the endowment is bounded below the discount rate, then the infinite future is irrelevant and bubbles of the sort just analyzed cannot occur. This is so because the loss of utility from future consumption is \((1 + g)x_0/(r - g) + b_0\), not \((1 + g)x_0/(r - g)\) as in the preceding case.

6.2 Sequential Markets

Even when trading restrictions prevent agents from directly arbitraging away bubbles, there exists an argument that can be used in some settings incorporating sequential markets to assure nonexistence of bubbles. Miguel Santos and Michael Woodford [80] and Huang and Werner [49] showed that if (1) agents’ optimal portfolios do not have bubbles, and (2) the aggregate endowment has finite value, then there cannot exist bubbles on securities in positive net supply. The reason is that if agents’ optimal portfolios do not have bubbles, the value of each agent’s consumption equals the value of his endowment (including the fundamental value of his endowment of the

---

13 Bewley was concerned with whether price systems have what he called an “economic interpretation”. Similarly, Stokey and Lucas [88], Chapter 15, analyzed conditions under which price systems have a dot-product representation, presuming also that the case in which the dot-product representation fails is uninterpretable.
security) plus the bubble on his endowment of the security. Summing over agents and invoking the assumption that the value of the aggregate endowment is finite, Walras’ Law implies that the aggregate value of the bubble on the security is zero. If the net supply of the security is strictly positive, it follows that the bubble itself must equal zero.

There exists a more intuitive way to see the connection between the value of the aggregate endowment and the existence of bubbles. Economies in which the aggregate endowment has finite value are economies in which the net interest rate is positive (or, in growing economies, greater than the rate of output growth). If there were a bubble on some security, the value of that security would eventually exceed the aggregate endowment of the economy. Transferring the security from generation to generation would imply negative consumption, invalidating the assumed solution path.\(^{14}\)

Example 6: Consider a deterministic overlapping generations model. Generation \( t \) \((t = 0, 1, 2, \ldots)\) is alive at dates \( t \) and \( t + 1 \) and has utility function \( \ln(x^t_t) + \ln(x^t_{t+1}) \), where \( x^t_{t+1} \) is the consumption of generation \( t \) at date \( t + 1 \). Each generation has an endowment of one unit of consumption when young and two units when old. The 0-th generation also has an endowment of money. In equilibrium each generation divides the value of its endowment equally between consumption when young and consumption when old, leading to

\[
x^t_t = \frac{p_t + 2p_{t+1}}{2p_t} \tag{25}
\]

\[
x^t_{t+1} = \frac{p_t + 2p_{t+1}}{2p_{t+1}}. \tag{26}
\]

Here \( p_t \) is the price of consumption at date \( t \) relative to an abstract numeraire. Combining these demand functions with the feasibility condition \( x^t_{t+1} + x^{t+1}_{t+1} = 3 \) results in the difference equation

\[
p_{t+2} = \frac{3p_{t+1} - p_t}{2}, \tag{27}
\]

which must be satisfied on any equilibrium path.

The restriction \( p_0 = 1 \) may be imposed as a numeraire choice, while \( p_1 \) indexes the equilibrium paths. Choosing \( p_1 = 1/2 \) results in the

\(^{14}\)This argument appears to depend on the restriction that only positive consumption is admissible, as is necessarily the case with the logarithmic utility that is usually specified in analyses of overlapping generations models. In fact positivity is not needed: even with utility functions that take negative as well as positive arguments, such as negative exponential utility, solution paths for which the assumed initial security prices exceeds its fundamental value fail in finite time. See LeRoy [59].
autarky allocation, so there is no bubble. Prices are given by $p_t = 2^{-t}$, so the aggregate endowment has finite value. Setting $p_1 = 0.48$ results in a bubble; generation 0 can consume 2.0417 when old instead of the endowment allocation of 2. Under the price system generated by (27), the value of the aggregate endowment remains finite.

<table>
<thead>
<tr>
<th>Date</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2.0417</td>
<td>2.0909</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.9583</td>
<td>2.2222</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.9091</td>
<td>2.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.7778</td>
<td>0.2</td>
<td>3.6667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.6667</td>
</tr>
</tbody>
</table>

As the accompanying table shows, this bubble increases in value, so the young of successive generations consume less and less and the old more and more. At date 5 consumption of the young becomes negative, invalidating the equilibrium path.

Examples 1 and 2 are consistent with the result just presented: in both cases the conditions for nonexistence of bubbles are not satisfied. In Example 1 the security with a bubble is in zero net supply, so existence of the bubble does not increase aggregate wealth. In Example 2, money can be valued only in overlapping generations equilibria for which the aggregate endowment has infinite value.

Now we are in a position to evaluate the contention that bubbles can be excluded on theoretical grounds under relatively weak assumptions. Andrew B. Abel et al. [1] argued that empirically security returns exceed the growth rate of the economy, implying that the aggregate endowment has finite value and that therefore, by the argument just given, bubbles cannot exist. This argument is formally correct, but it is entirely unconvincing as a proposition about the real world. The argument illustrates a recent tendency to appeal uncritically to rational expectations arguments, and also to rely, again uncritically, on an extreme reading of Milton Friedman’s positivist defense of optimization [30].

Let us deconstruct the argument that bubbles cannot exist (on securities in positive net supply). As noted, this argument presumes that agents can calculate the trajectory of the economy into the arbitrarily distant future. Representing agents as optimizing in an environment that they understand is most plausible in repetitive situations. In the context of bubbles, however, we are assuming that agents somehow come to understand the meaning of an event—model failure in the future along some

\[\text{Of course, it can be replied to this that the argument depends on agents acting as if they are making the calculation, not on the assumption that they actually do so. For example, following Friedman, it is not necessary that a skilled billiards player actually know the laws of mechanics, but just that he be able to play billiards as if he did. If he did not do so, he would not be a skilled billiards player.}\]
equilibrium paths—that by definition they have never experienced. This reasoning, which seems altogether implausible, goes unquestioned.

The same issue has come up in capital theory. In the 1960s economists studied efficient paths of capital accumulation in the presence of many capital goods, focusing on whether these paths can be implemented as competitive equilibria (the best introduction to this literature is Edwin Burmeister and A. Rodney Dobell [16]). Under conditions, Pareto-optimal equilibria display a turnpike property: from any initial condition the economy approaches that steady-state path on which the economy grows most quickly. In finite-time settings with a distant horizon the economy spends most of its time in the vicinity of the turnpike, and then exits the turnpike in order to satisfy an assumed terminal condition. In infinite time the economy converges toward the steady-state growth path. It follows that only one set of initial capital goods prices generates a path—that which converges to the steady state—that is feasible in infinite time. The other paths all fail in finite time.

Karl Shell and Joseph E. Stiglitz [82], writing before it was customary to appeal routinely to rational expectations, saw no reason to conclude from this that the nonconvergent paths were infeasible. Given the absence of the futures markets that would have revealed the impending crisis, they assumed that agents had no way to foresee future problems. It was only later that economists came to rely on the infinite-time feasibility of equilibrium paths to eliminate indeterminacy in dynamic economic models.16

The apparent mispricing of Royal Dutch/Shell shares constitutes evidence against the presumption that the Åbel et al. argument can be applied to the real world to eliminate bubbles.

Example 7: Two firms, Royal Dutch and Shell, divide a common pool of earnings in preassigned ratios. Despite this, the stocks of these firms typically trade at prices that diverge from the indicated ratio, sometimes by a wide margin (Kenneth J. Froot and Emil Dabora [31]). Since the firms have fundamental values that are proportional to the ratio in which they divide earnings, the mispricing implies that at least one of the firms has a bubble. Exploiting the bubble by buying the cheaper stock and selling short the more expensive one will produce profits if the mispricing disappears. However, the price gap might instead widen, producing losses, and the existence of trading restrictions might force arbitrageurs to close out their positions at a loss. Thus the transaction is not an arbitrage, and the mispricing is not eliminated.17 (Further, it follows that applying the term “mispricing” in this context is a misnomer).

---

16 These issues were discussed in Stiglitz’s [87] introduction to the 1990 Journal of Economic Perspectives symposium on bubbles, a paper that, although not widely cited, repays careful study.

17 As a digression, it is worth noting that the existence of limits to arbitrage has come to play a central role in supporting assertions that financial markets are inefficient (Shleifer [85], Nicholas Barberis and Thaler [7]). This is a complete inversion of logic. The existence of limits to arbitrage
7 Conclusion

We suggested above that characterizing asset price bubbles as manifestations of irrationality is unsatisfactory on methodological grounds. From the discussion of the preceding section it appears that difficulties remain even if obvious forms of irrationality are excluded. Unswerving adherence to neoclassical lines of reasoning produces an argument against bubbles that, although logically airtight, is simply not plausible. It is a testament to economists’ capacity for abstraction that they accept without question that an intricate theoretical argument against bubbles has somehow migrated from the pages of *Econometrica* to the floor of the New York Stock Exchange.

From within the neoclassical paradigm there is no obvious way to derail the chain of reasoning that excludes bubbles. We are, in effect, proposing thinking about bubbles in a neoclassical setting, but breaking off the analysis arbitrarily when doing so leads to conclusions that seem implausible. The problems with doing so are obvious: how does one write down formal models in such a setting? How does one know where to draw the line? Which conclusions from neoclassical analysis are to be accepted?

As an example, one curious implication of analyzing bubbles in the simplest overlapping generations setting under rational expectations is that they are welfare-improving: equilibria with big bubbles Pareto-dominate equilibria with little bubbles, and these Pareto-dominate equilibria with no bubbles. Is this outcome to be taken seriously as a proposition about the real world?

We have no good answers to these problems. However, even without such answers, the discussion here suggests several conclusions and indicates some directions for future research:

1. Assertions that theoretical and empirical evidence argue strongly against the existence of bubbles must be rejected. It is true that there exist theoretical arguments against bubbles, but we have suggested that these are implausible. Similarly, it is true that the empirical evidence of price and return volatility linked above with bubbles might have some other cause. That, however, does not constitute a showing that empirical evidence rejects bubbles.

2. Despite the foregoing argument, there is no prospect that bubbles—at least rational bubbles—can dispose of all of the major anomalies of asset pricing. The equity premium puzzle of Mehra and Prescott, for example, deals with
returns, not prices. Therefore existence of bubbles does nothing, at least in any
direct sense, to invalidate the Mehra-Prescott result that the equity premium
is excessive.

3. Demonstrating that models without bubbles do not generate as much price
and return volatility as the data appear to exhibit does not imply that models
with bubbles can do so. The link between bubbles and volatility may involve
sunspots (models in which equilibrium values of endogenous variables depend on
extrinsic uncertainty). Sunspots, although conceptually different from bubbles,
can occur under similar circumstances (non-Pareto-optimal equilibria). The
link between bubbles and sunspots and the connection between these and price
volatility have not yet been adequately investigated.

4. If bubbles and sunspots are part of the explanation of price and return volatility,
it follows that asset price fluctuations are not necessarily traceable to changes in
expected cash flows and discount factors. Thus an important piece of empirical
evidence that is often interpreted as rejecting capital market efficiency may not
necessarily support this interpretation.

5. As an corollary to the preceding point, from the rational bubbles perspective the
debate about capital market efficiency looks increasingly like a dispute about
whether the glass is half full or half empty. This is particular clearly in the
recent Journal of Economic Perspectives exchange between Burton Malkiel [68]
and Shiller [84]. Both describe a similar environment, one in which, in our
terms, there exist rational bubbles and sunspots. Malkiel concluded that even
though prices often depart significantly from fundamentals, markets are efficient
because the implied trading rules are only marginally profitable, and then not
for long. Shiller, on the other hand, associates market efficiency with the propo-
sition that security price movements are traceable to corresponding changes in
fundamentals, not with nonexistence of profitable trading rules. Given this
definition, he concluded that markets are inefficient.

References

Zeckhauser. Assessing dynamic efficiency: Theory and evidence. Review of


32


