

Fender, John

Working Paper

Towards a general theory of the stock market

Birmingham Business School Discussion Paper Series, No. 2016-03

Provided in Cooperation with:

Birmingham Business School, University of Birmingham

Suggested Citation: Fender, John (2016) : Towards a general theory of the stock market, Birmingham Business School Discussion Paper Series, No. 2016-03, University of Birmingham, Birmingham Business School, Birmingham,
<http://epapers.bham.ac.uk/2102/>

This Version is available at:

<https://hdl.handle.net/10419/202667>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-sa/2.5/>



UNIVERSITY OF
BIRMINGHAM

BIRMINGHAM
BUSINESS
SCHOOL

Birmingham Business School Discussion Paper Series

Towards a general theory of the stock market

John Fender

2016-03



This discussion paper is copyright of the University and the author. In addition, parts of the paper may feature content whose copyright is owned by a third party, but which has been used either by permission or under the Fair Dealing provisions. The intellectual property rights in respect of this work are as defined by the terms of any licence that is attached to the paper. Where no licence is associated with the work, any subsequent use is subject to the terms of The Copyright Designs and Patents Act 1988 (or as modified by any successor legislation).

Any reproduction of the whole or part of this paper must be in accordance with the licence or the Act (whichever is applicable) and must be properly acknowledged. For non-commercial research and for private study purposes, copies of the paper may be made/distributed and quotations used with due attribution. Commercial distribution or reproduction in any format is prohibited without the permission of the copyright holders.

TOWARDS A GENERAL THEORY OF THE STOCK MARKET

John Fender (University of Birmingham)

October 2015

Abstract: Although there are many stock market anomalies which the Efficient Market Hypothesis (EMH) finds difficult to explain, it also has its strengths, and so far no alternative hypothesis has been developed which can explain what the EMH explains but which can also do a better job in explaining the phenomena with which it struggles. It is argued that the way forward is to postulate that the stock market can be in one of three states: a fundamental state, in which share prices are determined as in the EMH, a bubble or bull market state, in which share prices are above their fundamental levels but continue to rise because asset holders expect to sell the shares at even higher prices in the future, and a bear market state, in which shares are held exclusively by 'irrational' agents and rational agents cannot exploit the overvaluation because of short-selling constraints. It is also argued that heterogeneous rational expectations may help explain some features of stock market behaviour.

Keywords: efficient market hypothesis, rational expectations, bubbles, bear markets, short-selling constraints.

JEL Classification Code: G1.

Acknowledgements: Earlier versions of this paper were presented at a University of Birmingham internal workshop, at a conference at Lancaster University in Honour of Robert Rothschild, at the Research Day at Aston University in 2015 and at the 2015 Money, Macro and Finance Research Group conference in Cardiff. The author is grateful to a number of participants who made helpful suggestions.

The contents of this paper should be considered preliminary. Please do not quote without the author's permission. Comments are very welcome.

Professor John Fender
Department of Economics
University of Birmingham
Edgbaston
Birmingham B15 2TT
United Kingdom

Phone: (44) 121 414 6644
Fax: (44) 121 414 7377
Email: J.Fender@bham.ac.uk

1. Introduction

The 2013 award of the Nobel Prize in Economics to both Eugene Fama and Robert Shiller seemed a puzzle to many, since they are associated with very different and seemingly incompatible views of how the stock market works.¹ Fama, in particular, is regarded as the leading academic proponent of the Efficient Markets Hypothesis (EMH) and Shiller is the leading academic critic of the hypothesis. Surely they both cannot be right?

The purpose of this paper is to argue that Fama and Shiller are both (largely) right, and to outline an approach to explaining the behaviour of the stock market which can reconcile their seemingly divergent views. According to this approach share prices can be determined as postulated by the EMH, modified to incorporate heterogeneous rational expectations and transactions costs. However, the stock market can be in one of two other states or regimes: a bubble or bull market state, in which share prices are above their fundamental values but rational agents may hold the shares because they expect share prices will rise further, and a bear market state in which shares are overvalued and held exclusively by irrational agents; rational agents cannot exploit the overvaluation because of short-selling constraints. One important consideration is that agents in each state may need to take into account the implications of the fact that the state may change – for example, the fundamental state may change into a bubble, or a bubble might collapse and the market may enter a bear market state.

The paper proceeds by first of all outlining the EMH and its strengths. The anomalies are then discussed, and much of the remainder of the paper is devoted to developing the

¹ There was a third Nobel laureate in 2013, Lars Peter Hansen, whose work is less relevant for the topic of this paper. For an evaluation of the three 2013 laureates' contributions, which emphasises their similarities rather than their differences, see Campbell (2014).

alternative hypothesis and discussing how it might explain the behaviour of the stock market.

2. The Efficient Market Hypothesis

Since the EMH is central to the understanding of stock markets, it is worthwhile describing its essential features, its underlying assumptions and some of the evidence in its favour.

To start, we might take the following fundamental equation, (which we describe, following the literature, as the standard efficient market or no arbitrage condition):

$$P_t = E[P_{t+1} + D_{t+1}] / (1 + r). \quad (1)$$

Notation is standard. P_t is the price of a share at time t and D_t is the dividend it pays. The rate at which investors discount future returns is given by r . For simplicity this is assumed to be constant, but there is no problem (at least in principle) in making it time varying. The equation states that the current price of a share equals the expected price next period plus the dividend expected to be received next period, appropriately discounted.² We can interpret r as the expected return investors must be offered to hold a share, so equation (1) essentially states that it is expected that this share generates this return. If the equation fails to hold, then investors will make purchases or sales which will restore equality. For example, suppose the expected future price plus dividend (i.e. $P_{t+1} + D_{t+1}$) is greater than $(1 + r)P_t$. Then investors can get a higher expected return by holding this share than from holding other shares, so we would expect investors would sell these other shares to buy the share in question, and this process will continue until equality is restored.

² It is assumed that the price at time t is the 'ex dividend' price as far as any dividend payable at that time is concerned.

Equation (1) should hold if all investors are rational and have identical expectations and also if there are no transactions costs. Note that these assumptions are sufficient and not necessary. A point often made by proponents of the EMH is that it is by no means necessary for equation (1) to hold that *all* investors be rational. It is only necessary that ‘*enough*’ investors are rational. Suppose there are many non-rational investors, and, in the absence of rational investors, their share buying and selling behaviour would produce an expected return on the share of less than r . Then rational investors should sell the share in question (they might sell it short if they do not hold the share), and in doing so the share price falls until equation (1) is re-established.

The assumption of zero transactions costs is probably not that crucial. There are many investors for whom the transactions costs of buying and selling shares are fairly small, so we would expect equation (1) to hold approximately even with transactions costs.

However, equation (1) is not the EMH. By iteration and repeated substitution (and applying the law of ‘iterated expectations’), we can obtain the following expression for share prices at time 0:

$$P_0 = \lim_{T \rightarrow \infty} \frac{E[P_T]}{(1+r)^T} + \sum_{i=1}^{i=\infty} \frac{E[D_i]}{(1+r)^i}. \quad (2)$$

This states that the price of a share at time t is the sum of two components. The first term – the expected present value of the discounted price as time goes to infinity – might be described as a ‘bubble’ term (so we define the bubble term as $B_0 \equiv \lim_{T \rightarrow \infty} \frac{E[P_T]}{(1+r)^T}$). The second term is the discounted present value of the expected future stream of dividends and may be described as the ‘fundamental value’ of the share. The EMH may then be stated as the hypothesis that the price of a share equals its fundamental value:

$$P_0 = \sum_{i=1}^{t=\infty} \left(\frac{E[D_i]}{(1+r)^i} \right) \equiv F_0. \quad (3)$$

This follows from equation (2) with the additional assumption that the bubble term is zero. This is how we will interpret the EMH. The essential debate is over whether this equation is a basically correct theory about how share prices are determined. Note that the EMH depends on two components. We need to combine an assumption that the efficient market condition holds with the assumption that the bubble term is zero. We will discuss this latter assumption later in the paper.

We will be brief in discussing the evidence in favour of the hypothesis. There is a huge amount of relevant empirical evidence. The bulk of empirical work in the 1960s and 1970s was supportive: e.g. ‘there is no other proposition in economics which has more solid empirical evidence supporting it than the Efficient Markets Hypothesis’ (Jensen, 1978, p. 95). More recent work has been more sceptical; we will discuss some of this work later. However, it is probably fair to say that there are two predictions of the hypothesis which have survived fairly although perhaps not completely intact. These are the results that

- (1) Stock prices react to news, and
- (2) It is impossible to beat the market systematically.

This first prediction seems trite, almost trivial. Every day, there are reports in the financial press of news affecting companies and how their shares prices react. That share prices react to news is not in doubt. Whether they react exactly to the extent that the EMH predicts is another question. This is a difficult question to answer as it is not easy to obtain a good estimate of a firm’s fundamental value and how it changes. The second prediction seems to hold as well. There are many studies showing that so-called ‘experts’ do not, on average

beat the market (see, for example, Malkiel, 2003, pp. 76 – 80). Consider, for example, the following well-known passage from Richard Roll, an academic financial economist who is also a portfolio manager:

‘I have personally tried to invest money, my client’s money and my own, in every single anomaly and predictive device that academics have dreamed up...I have attempted to exploit the so-called year-end anomalies and a whole variety of strategies supposedly documented by academic research. *And I have yet to make a nickel on any of these supposed market inefficiencies...*a true market *inefficiency* ought to be an exploitable opportunity. If there’s nothing investors can exploit in a systematic way, time in and time out, then it’s very hard to say that information is not being properly incorporated into stock prices’ (quoted in Malkiel, 2003, p. 72).

So we will consider it a basic requirement of any alternative theory of the stock market that it is consistent with the predictions of the EMH that stock prices react to news and that it is impossible (or at least extremely difficult) to beat the market systematically.

3. Empirical Evidence against the EMH

However, since the early 1980s, there have been a number of empirical challenges to the EMH, and there is now a long list of so-called ‘anomalies’, of which we would mention the following:

(1) Shares are sometimes massively overvalued, in the sense that they far exceed any reasonable estimate of their fundamental value. There are examples from the internet bubble which ended in 2000. Appendix I gives one such example. Papers which argue for this conclusion include Ofek and Wilson (2002).

(2) Shiller (1981) argued that stock prices are far more volatile than the EMH can justify ('excess volatility'). There has been much debate about these conclusions. Shiller (2014) provides a fairly up-to-date account of the literature and results on excess volatility. There seems to be a consensus now that Shiller's contention is basically correct – it does not seem to be something that the proponents of the EMH have effectively challenged.

(3) Bubbles in share prices sometimes occur. A 'bubble' occurs when a share price is above its fundamental value, but rises in price since it is bought by agents who expect to resell it at a yet higher price. So equations (1) and (2) above hold in a bubble, but not equation (3). Plausible examples are the South Sea Bubble and Dutch Tulip mania. (For a list of famous bubbles, see pp. 170 – 1 of Shleifer, 2000.) A more recent example is the dot-com bubble that ended in March 2000. But it is difficult to establish that these are bubbles; after all, large increases in share prices are not incompatible with the EMH – they can be explained easily if there is a large change in fundamentals.

There is a massive literature testing for the occurrence of bubbles, which it is impossible to discuss at any length here. There are also historical analyses of various purported bubble episodes (e.g. Garber, 1990), which argue that they were no such thing. But there are other analyses of such episodes which reach different conclusions. For example: 'the qualitative evidence assembled in this paper favors the view that a bubble was present in the 1929 market' (White, 1990, p. 67). Without wishing to make a judgment about whether specific market episodes were or were not bubbles, we regard the contention that stock market bubbles sometimes exist as highly plausible, and how and why they occur is something that any satisfactory theory of the stock market needs to explain. For a recent survey of the literature on bubbles, see Scherbina and Schlusche (2014).

(4) Crashes in share prices happen and large downward movements in share prices are much more common than large upward movements. (See, for example, Hong and Stein, 2003). While large falls in stock prices are not, of course, incompatible with the EMH, the EMH would seem to require that such falls are due to large negative pieces of news, and many share price crashes do not seem to be accompanied by such news.

(5) There are the so-called calendar anomalies (see, for example, chapter 18 of Siegel, 2002). One example is provided by the fact that the return on Mondays over the period 1946 – 2001 on the Dow Jones was -0.0689% whereas that on Fridays was 0.0734%; the biggest return is on the last day before Christmas, of 0.3291%. (See Table 18-2 on p. 311 of Siegel, 2002.) The most famous calendar anomaly is of course the January effect; for some relatively recent evidence that the effect has persisted into the twenty first century, see Haug and Hirschey (2006).

(6) There is evidence that ‘momentum’ strategies, whereby shares that have recently risen (fallen) in price are bought (sold) do generate excess returns. For example, Jegadeesh and Titman (2001) present evidence that strategies that buy stocks that have done well over a period of 3 to 12 months and sell stocks that have done badly over the same period tend to generate profits of about one percent per month over the following year. Fama has described the momentum result as ‘the biggest challenge to market efficiency’ (Fama, 2014, p. 1480). Actually, these studies are ‘ex post’ studies that show that if one had used these strategies in the period under consideration, one would have made excess returns. It does not follow that one can use these strategies to make excess returns in the future – it might well be that enough people start using the techniques and in doing so eliminate the opportunity to make excess returns.

(7) Long run 'mean reversion' in stock prices may occur. If share prices follow a random walk, which is an implication of the EMH with a constant discount rate, then the level of the stock market will be irrelevant to its subsequent movement. So if share prices are high, this has no implications at all for whether they are likely to continue to rise or fall. However, there is some evidence for 'mean reversion' in share prices (the classic reference is Poterba and Summers, 1988). So if share prices are well above their long-run average, they are likely to move back toward this average, and so forth.

(8) Shleifer (1986) conducted an ingenious test of the EMH, by examining the behaviour of shares which were included in, or excluded from, certain stock prices indices. When an announcement is made that a certain share is to be included in a stock market index, this may induce buying by index funds, which hold shares not on the basis of their expected returns, but in accordance with their weight in a certain index. Shleifer found that announcements that certain shares were to be included in, or excluded from, a stock market index did have significant effects on their prices. This seems to contradict the EMH, according to which a decision by an agent to purchase a share should not change its price in the absence of a change in its fundamental value – if such a purchase does tend to raise a share price above its fundamental value, this will be offset by share sales by rational agents which bring the share price back to its fundamental level.

(9) There is some evidence of both underreaction and overreaction to news announcements. (See pp. 114 - 120 of Shleifer, 2000, for evidence of underreaction; the following pp. 120 – 127 contains evidence of overreaction.) It might be asked how the stock market can exhibit both overreaction and underreaction; it would seem to be possible if the reactions take place over different horizons – it may be that underreaction takes place in the few months

after (say) an earnings announcement, whereas overreaction takes place over several years following a series of announcements of good or bad news.

This is by no means a comprehensive list of the purported empirical failings of the EMH. Nevertheless, not all adherents of the EMH have been persuaded. For most of the above anomalies, explanations have been put forward which are compatible with the EMH. For example, large increases in stock prices can be explained by a big improvement in fundamentals or by a significant fall in the discount rate. We will briefly review the question whether the EMH is compatible with these seeming anomalies below.

4. Investment Advice, the Views of Market Practitioners and the Behaviour of Investors

If the stock market does indeed behave (approximately) as the EMH predicts, but there are transactions costs in buying and selling shares, then the implications for investor behaviour are quite clear cut. A risk-averse investor should buy and hold a diversified portfolio of shares. She should only trade to re-balance her portfolio or to change the overall size of her portfolio. So, if she owns a share that rises considerably in price, she may become overweight in that share, so considerations of risk management should induce her to sell part of her holding of the share. If she owns a share that falls in price, she should emphatically not sell it – if anything, she should buy more of it, since she may be underweight in it. But this differs significantly from the advice often given to investors. For example, many investment advisors recommend a ‘stop-loss’ rule, whereby a share is automatically sold if it falls more than (say) 20% below a previous peak. Such a rule makes absolutely no sense if one believes in the EMH. The fact that a share price may have fallen considerably has no implications for the future movement of that share, and does not mean

that its expected return has fallen. If anything, one should buy more of a share that has fallen in price, in order to re-balance one's portfolio.³

There are many other ways in which stock market professionals deviate from what the EMH would predict. For example, shares are often given 'buy' or 'sell' recommendations by analysts. If the EMH holds, then all shares have the same expected return adjusted for risk, so such recommendations are pointless. (What would be valuable would be information on the risk properties of various shares – more specifically, information on how returns on different shares are correlated with each other, which would be helpful in deciding on the structure of a portfolio that minimises risk for any given expected return, but such information is not typically given.)

Much discussion of stock markets and investment advice presupposes a 'non EMH' framework. For example, there is considerable discussion of 'bull' and 'bear' markets. But if the EMH holds, it is difficult to see how there can be a bear or bull market. A bull market is one in which shares are persistently rising; if the market rises because of the realisation of a few pieces of good news, then this has no implications at all for the future movement of the market – believing that the market will rise further because it has risen in the recent past is hence analogous to believing that if a coin is tossed a number of times and usually comes up 'heads' then such an outcome is more likely in the future.

A further point is that turnover in the stock market seems much greater than what would be consistent with investors behaving in accordance with the EMH, according to

³ For an attempt by the author to show the absurdity of a stop-loss rule, see Fender (2001). These arguments did not meet with universal approval, as a perusal of the letters columns of the *Investors Chronicle* in the following few weeks will reveal.

which investors should sell shares only if they need to re-balance their portfolios or if they need to reduce the overall size of their portfolio (perhaps to finance consumption).

Perhaps practitioners are just ignorant. But another possibility, which is at least worth considering, is that their views may be more reasonable if the stock market operates in a way less consistent with the EMH. Indeed, the approach we present in this paper may be more compatible with the views of many practitioners, something we regard as mildly supportive of our approach.

5. Theoretical Arguments Relating to the EMH

The EMH will hold if agents are homogenous and rational, if there are no transactions costs and if bubbles can be ruled out. Of course these assumptions do not hold exactly, and economists have devoted considerable effort into exploring whether and how the hypothesis changes if the assumptions are relaxed. In this section we discuss some relevant contributions that have been made.

One powerful argument was that of Grossman and Stiglitz (1980) who established that there cannot be a perfectly informationally efficient market if there are costs of gathering information. Suppose that the EMH holds and shares follow a random walk, but also that it is costly to gather information about the fundamental values of shares. Then the return to gathering such information will be zero so no information will be gathered. But this means there will be no one to identify, and step in to rectify, divergences from market efficiency, so there is no reason for the market to be efficient. Presumably share prices need to diverge from their fundamental values sometimes; such divergences can be exploited by those agents who gather information in order to exploit asset inefficiencies. So, it would seem, the stock market, and other markets, cannot be perfectly efficient. But this does not

prevent the markets being approximately efficient, with small divergences from efficiency generating the incentives for some agents to gather enough information – and exploit it - to make the market behave in this way.

Another argument was that of Miller (1977), who pointed out that if there are short-selling constraints, there may be no way in which rational agents can exploit stock market overvaluation. Perhaps share prices have been inflated by the purchases of a few zealots. If rational agents do not hold the shares, the only way they can exploit the overvaluation is to sell the shares short, but short-selling constraints may prevent them from doing this.

The above arguments are arguments for thinking that the efficient market equation (equation (1)) may not always hold. As far as the second component of the EMH is concerned (the assumption that the bubble term is zero), there has been considerable discussion in the literature of whether rational bubbles are possible. If a share is overvalued, then it still may be held by rational investors if they expect they will be able to sell it to investors in the future at a still higher price. But these investors will only, rationally, be willing to buy the share at this higher price if they expect they will be able to sell it at a yet higher price sometime in the future, and so on *ad infinitum* (the ‘greater fool’ hypothesis). Is this compatible with rationality? A number of papers have explored this issue rigorously (e.g. Tirole, 1982 and 1985, Santos and Woodford, 1997). It turns out that in many economies, bubbles can be ruled out. The basic argument is that ‘the existence of a bubble would require asymptotic growth in the value of the asset in question, and hence asymptotic growth of the wealth of at least one of the households, at a rate inconsistent with optimization by that household’ (Santos and Woodford, *op. cit.*, p. 20). Further work has confirmed that the conditions under which rational asset pricing bubbles are possible are

fairly restrictive. Should we then conclude that bubbles are implausible? It might be pointed out that the type of rationality necessary for these results to go through is very strong. It is necessary that investors at time t believe that investors at all future times ($t + 1$, $t + 2$, $t + 3$,...) are rational. But this is not enough. It is also necessary that investors at any time believe that investors in the next time period believe that investors in the future will be rational, and so forth. An extremely demanding condition, analogous to common knowledge, is required. It can be seen how, if the condition fails, a bubble can emerge. Suppose investors at time t are rational and believe that investors at all future times are rational, but also believe that investors at some future time $t + m$ do not believe investors at time n (with $n > m$) will be rational. Specifically, suppose they believe that because of this non-rationality investors at time $t + m$ will believe that share prices at time $t + n$ will be twice their fundamental value. Then investors at time t believe investors at time $t + m$ will bid share prices up because of this anticipated overvaluation, and hence they likewise bid share prices up at time t for the very same reason. So it seems that the requirements on beliefs needed to rule out bubbles are very strong – it needs to be believed that there is no chance that asset prices in any future period will be overvalued, that there is no chance investors believe that investors at any time period in the future may not be rational, that there is no chance investors believe that future investors believe that investors in the even further future will be anything other than rational, and so forth.

6. Detour on Methodology (Kuhn, with influence from Popper and Quine)

Here an approach to explaining scientific revolutions pioneered by Thomas Kuhn is outlined (see Kuhn, 1970) which, although it was developed with the case of Einstein's overthrow of Newtonian physics in mind, will be argued to be particularly relevant in

explaining the current and recent state of academic work on the stock market. The fundamental concept in Kuhn's theory is that of a 'paradigm', which may be defined as a 'universally recognized scientific achievement that for a time provides model problems and solutions to a community of practitioners.' (Kuhn, 1970, p. viii). Newtonian physics was such a paradigm for several hundred years; it explained many phenomena and seemed to offer the hope of explaining many more.

Much of scientific activity is 'normal science', attempting to use the paradigm to explain phenomena that have not hitherto been explained using the paradigm, but which practitioners believe may be capable of being so explained. It may take a great deal of work to do this. Sometimes the relevant hypothesis does not seem to be confirmed but that does not lead to rejection of the hypothesis – it is (almost) impossible to refute a scientific hypothesis unambiguously, because testing scientific hypotheses invariably involves testing several hypotheses (i.e. one needs auxiliary hypotheses as well as the main hypothesis that is being tested), so that if the test fails, it is not clear what is responsible. If the test fails, a sensible approach **may** be to assume that one of the auxiliary hypotheses fails.

However, there may come a time when 'anomalies' accumulate. These are observations which seem to contradict the hypothesis, and which cannot be explained away despite persistent attempts. An example would be the period, starting in the late nineteenth century and which lasted until Einstein's special theory of relativity emerged in 1905, when Newtonian physics seemed unable to explain many anomalies but there was no plausible alternative explanation. In such a period there is a 'crisis' in the science – the dominant paradigm is believed (by many) to be unsatisfactory, but there is no consensus on what should replace it.

Eventually, a new paradigm may appear; this needs to be capable of explaining what the old paradigm explained, but seems to offer a better chance of explaining what the old paradigm could not explain. In physics, this was Einsteinian physics, which was able to explain what Newtonian physics could but which could also explain - or at least offered the hope of explaining - some of the anomalous phenomena.

It should be clear how we might attempt to apply this approach to the theory of the stock market. The dominant paradigm is, of course, the EMH. When this emerged, it explained and clarified numerous observations, and provided hope of explaining many other unexplained phenomena. A massive amount of work in finance has been within the framework of the EMH paradigm. But it seems to be the case now that a fair number of anomalies have emerged which cannot be easily explained by the paradigm. So what happens now? According to Kuhn: ‘... all crises close in one of three ways. Sometimes normal science ultimately proves able to handle the crisis-provoking problem despite the despair of those who have seen it as the end of an existing paradigm. On other occasions, the problem resists even apparently radical new approaches...Or, finally...a crisis may end with the emergence of a new candidate for paradigm and with the ensuing battle over its acceptance’ (*op. cit.*, p. 84).

One possibility is that the current crisis in the theory of the stock market could be resolved in favour of the EMH. But given the number and persistence of the anomalies, this seems unlikely. It is perhaps possible that no remedy will emerge – this is Kuhn’s second way. The third possibility is that a new paradigm emerges which eventually supplants the EMH. Do we have such a candidate at the moment? Some may suggest behavioural finance. But it seems difficult to disagree with Fama who writes: ‘the behavioral literature has not

put forth a full blown model for prices and returns that can be tested and potentially rejected – the acid test for any model proposed as a replacement for another model.’ (Fama, 2014, p. 1477.) There does not seem to be a behavioural theory of the stock market that can explain why it is impossible to systematically beat the market, why stock prices react to news and why bubbles and crashes might occur, for example. To qualify as a candidate to supplant the dominant paradigm, the alternative hypothesis must account for what the previous paradigm can explain, and also offer hope of resolving the anomalies. In the rest of this paper we put forward some suggestions of how such a theory might be developed. But first, in the next section, we discuss whether it is possible for the EMH to be extended in any plausible way to explain the anomalies.

7. Can the EMH Explain the Anomalies?

Considerable effort and ingenuity have been expended by adherents of the EMH to explain the empirical anomalies listed in Section 3. We do not discuss these attempts at any length. However, we would mention that believers in the EMH seem to acknowledge that the existence of transactions costs means the market will not be perfectly efficient, and that small deviations from market efficiency are required to generate incentives for agents to gather information to exploit market inefficiencies. If this is the case, then there may be a ‘band’ about the fundamental price of a share within which the share price can fluctuate without there being exploitable excess returns. It may be possible to explain some anomalies in this way. Take for example, the calendar anomalies mentioned above – these anomalies suggest fairly small excess returns from strategies such as (say) buying just before a holiday and selling afterwards, and it is quite possible that these excess returns are outweighed by the transactions costs of implementing the strategy.

Calculation costs may also be significant. It may not be a trivial task, by any means, to work out a firm's fundamental value. These are costs that presumably are borne by the traders who intervene to ensure market efficiency. They may need to study a large number of firms before they find one that is sufficiently mispriced so as to offer an opportunity of excess returns. To compensate traders for these costs, there presumably need to be some further deviations from market efficiency which they can exploit. Also, a trader may need to make a large purchase or sale of shares to gain from a mispricing of a firm's shares sufficient to outweigh these costs, which might be largely independent of the number of shares transacted. It may be difficult to make these purchases without changing the share price (but this of course depends, *inter alia*, on the size of the firm). For example, suppose that a trader needs to make an excess profit of £10,000 each trade to compensate for the costs of investigating this particular trade and also the costs of investigating shares he decides not to buy or sell. Then if he is contemplating buying 100,000 shares, he will only do so if he can obtain them at a price at least 10p below their fundamental value. If such a purchase raises the share price by, say, 5p, then he will only make the purchase if the initial price of the shares is at least 15p below their fundamental value. So this may be another reason why there may be a 'band' around the fundamental value of a share, within which its price can fluctuate without inducing purchases or sales by traders expecting to make excess returns. Also, the band, if it exists for this reason, may be considerably greater for small market capitalisation firms than for large ones.

It seems usually assumed in the exposition of the EMH that rational investors are all the same, so have the same discount rate and same expectations. But there is no need for this to be the case. Surely agents could have different discount rates, perhaps because of

differences in risk aversion? Also, it might be suggested that not all rational agents have the same expectations. We discuss the plausibility of this assumption in Appendix II.

If rational agents do indeed have heterogeneous expectations, then the EMH needs reformulating, since there is now no such thing as 'the' fundamental value of a share – each agent will have his own estimate of the fundamental value of a share. So the EMH should state that the value of a share equals its fundamental value as estimated by the marginal trader.

If there is such heterogeneity amongst rational agents, one or two of the anomalies can be explained. Such heterogeneity can give rise to downward sloping demand curves for stocks, so that Shleifer's evidence about the effect of shares being included in stock market indices can be explained. If index funds start buying a large quantity of a particular share for this reason, then the shareholders with the most pessimistic expectations about its prospects may have already sold their holdings, so that the purchasers need to offer a higher price to induce holders of the shares with more optimistic expectations to part with their holdings.

How much share prices change in response to an exogenous change in supply is of course an empirical question. One might conjecture that for some shares, held mainly by a few traders who derive their information from the same sources, the demand curve might be close to horizontal. However, for some other shares, perhaps held by traders with very different information sets and beliefs, the price elasticity might be much greater.

So it may be possible for the EMH to explain some of the anomalies if it is extended to incorporate transactions costs, calculation costs and also heterogeneous rational agents. But it seems very difficult to explain all the anomalies. The excess volatility argument, the

extreme overvaluation of certain shares and momentum effects seem very difficult to explain even with these extensions of the hypothesis. It seems reasonable, then, to consider whether there are other ways of explaining these anomalies, and this is what we intend to do in the remainder of this paper.

8. Foundations of a General Theory of the Stock Market: Irrational Agents and Short-Selling Constraints

So how do we explain the anomalies, but preserve the insights of the EMH? We need a 'General Theory', one which explains what the EMH does not explain as well as what it does explain.

Let us consider the case of the incredibly overvalued share price (of Baltimore Technologies) considered in Appendix I below, and ask how its shares could be so valued. There are two possibilities. One is that there is a bubble – the share price is overvalued because there are expectations that it will be even more overvalued in the future. But suppose there is also no general expectation of a rise in the share price, so the bubble explanation does not work; could the overvaluation still exist? It could, provided that two conditions are met:

(1) The shares must be held by someone. However given that they are above their fundamental value and there is no expectation of a rise in their prices, they cannot be held by rational investors. Therefore they must be held by irrational investors. So the existence of irrational agents, often called 'noise traders', is necessary for such an overvaluation.

(2) It must be impossible for rational investors to exploit the overvaluation, in this case by short selling. There is much literature on short-selling constraints and agreement that such selling is generally costly and risky (see e.g. Jones and Lamont, 2002). It is sometimes

regarded as reasonable to assume that shorting is impossible and this is the approach we will take here – we will assume that short selling is either impossible or costly and rarely undertaken, and discuss some of its implications. (An assumption that short selling is impossible is fairly common in the literature now).

The basic idea behind this paper is that the key to developing a ‘General Theory of the Stock Market’ (GTSM) involves three crucial assumptions:

- (1) An assumption that there are a considerable number of irrational traders (as well as a large number of rational investors) active in the stock market.**
- (2) An assumption that short selling is impossible (or at least costly and fairly rare).**
- (3) An assumption that rational and irrational agents may interact in different ways in different circumstances.**

If this is done, we can explain how some shares can sometimes be overvalued. Note also that now the EMH (interpreted as the theory that stock prices are determined as in equation (3)) is not the same as the view that there cannot be exploitable excess returns. If a stock like Baltimore Technologies is overvalued by the EMH criterion, because it is held by irrational traders who have ‘fallen in love’ with the stock, and rational agents cannot exploit the overvaluation because of short-selling constraints, then there are no exploitable opportunities to make excess returns even though equation (3) does not hold. In this case, the following inequality would replace equation (1):

$$P_0 > E[P_1 + D_1] / (1 + r). \quad (4)$$

We now come to the basic idea behind the GTSM:

The stock market can be in one of three states. It can be in a ‘fundamental state’; this is where equations (1) and (3) hold. Secondly, it can be in a bubble or bull market state; this is

where equation (1) holds but equation (3) does not (instead share prices are above their fundamental values). Thirdly, it can be in a bear market state; this is where shares prices are above their fundamental values but equation (1) does not hold; instead inequality (4) is satisfied.

Much needs to be done to explain the behaviour of the stock market in each of the non-fundamental states (behaviour of the stock market in the fundamental state is much better understood than it is in the non-fundamental state, but there are still some puzzles.) Also, we need to be able to explain why and how one state may turn into another. In what follows, we discuss each of these states in turn.

9. The Bear Market

A bear market may arise after the collapse of a bubble; share prices may not fall all the way to their fundamental values because shares are bought by noise traders before they do so. (Note that this requires at least some noise traders to be price sensitive – i.e. they buy more shares as their prices fall.) How prices change in a bear market depends on the behaviour of irrational agents. Here the insights of behavioural finance might be particularly relevant. We might expect share prices to fall, on average, for a number of reasons:

1. There might be some short-selling pressure (an assumption of zero short selling is surely too extreme).
2. Irrational agents might come to believe that the market is overvalued, and sell. (Some agents may not be totally irrational, but may learn over time.)
3. Momentum strategies are quite important in practice, so the initial fall in prices might give rise to further selling pressure, further price falls, and so forth.

4. If overvalued shares are held by index funds, then as share prices fall, such shares might be expelled from the indices. The index funds therefore sell them and share prices hence decline further.

5. If some agents have bought shares on margin, a decline in share prices might prompt margin calls, and agents might need to sell shares for this reason.

We might expect fundamental values to rise over time; as share prices in a bear market can be expected to decline over time, eventually prices will return to fundamental values and we switch to a 'fundamental' regime. Share prices in a bear market might be expected to be particularly unpredictable (but not necessarily follow a random walk).

So far we have thought of the bear market as comprised entirely of irrational traders. However, if we admit the possibility of heterogeneous rational agents, we can think of the bear market as comprised of both irrational and rational agents, but where the marginal investor is irrational. This provides another reason why the demand curve for shares in a bear market is downward sloping (as the share price falls, it falls below its fundamental value as estimated by more and more of the most optimistic rational agents, who hence buy); reversion to the fundamental market occurs when the marginal trader is a rational agent.

There is also no reason why in a bear market share prices should not react to news, but this of course depends on irrational agents reacting to news. Whether they do, and if so the extent they do, may well depend on what type of irrational agent they are. This is something that needs further research.

10. The Bull Market (Or Bubble)

This is characterised by equation (2) where the first term on the RHS (the bubble term) is positive. But equation (1) continues to hold. There are a number of points that might be made:

1. Participation in a bubble is by no means irrational. It is rational to continue to hold shares as long as equation (1) holds, but this equation holds in a bubble. Evidence that some of the players in the South Sea bubble were rational is provided in Temin and Voth (2004). Similar evidence for the dot-com bubble between 1998 and 2000 is provided by Brunnermeier and Nagel (2004).

2. Whether a rational bubble is possible (i.e. a bubble with all agents rational, which means they have, inter alia, appropriate 'higher order' beliefs about the rationality of other agents, including agents in the far future) has been a controversial and much studied topic. As argued in Section 5, these rationality assumptions are very demanding (much more demanding than the rationality assumption required to support equation (1), for example, and even this assumption may fail occasionally). We therefore need to develop models of bubbles that relax such extreme rationality assumptions, but there are few such models in the literature. (Scheinkman and Xiong, 2003 and Abreu and Brunnermeier, 2003, are examples.) Developing tractable models of bubbles with rational and irrational agents would seem to be an important item on the research agenda.

3. We do not know how a bubble starts. A plausible story is as follows: there is an important innovation that raises fundamental values in at least one sector of the economy. Rational agents hence bid share prices up. Irrational agents extrapolate the rise in prices and buy,

hence raising share prices above their fundamental levels; rational agents anticipate this overvaluation and buy as well.

4. We also do not know how a bubble terminates. Some literature has stressed the need for a 'synchronising event' that brings an end to the bubble.⁴ The bubble may be kept going by agents believing that other agents believe that share prices will keep rising (there may be higher order beliefs involved as well); the bubble might start to collapse when some agents become sceptical that other agents believe that another group of agents believe that share prices will keep rising (or something similar), but getting the details right and modelling this in a plausible way may be very difficult indeed. It seems we are a long way away from having a satisfactory model of bubbles which would, in particular, explain how and when they start and how and when they end.

We might comment that if bubbles are ended by such synchronising events, then this would be an example of share prices reacting to news, but the movement in share prices might be very different from the change in the fundamental value of the shares. This is one way in which share prices in a bubble might react to news. A second way is that the size of the bubble might be affected by news. Whilst it is therefore perfectly compatible with the economy being in a bubble that share prices should react to news, there is virtually no empirical or theoretical work that explores how share prices react to news in a bubble – this is clearly something that needs further work.

5. In a stochastic bubble (see Blanchard and Watson, 1982, pp. 297 – 299), both the probability of a collapse and what happens when the bubble collapses are relevant in explaining the behaviour of a bubble. If a bear market is expected to result after the bubble

⁴ See for example, pp. 193 – 197 of Abreu and Brunnermeier (2003).

ends, then the level of share prices in the bear market may well affect what happens in the bubble. So if (for example) something happens that increases the size of index funds, this may mean that share prices are not expected to fall as much following the end of a bubble, hence strengthening the bubble.

6. We would conjecture that there are plenty of opportunities for behavioural finance to explain aspects of agents' behaviour in a bubble, and hence be very relevant in developing plausible models of bubbles.

11. The Fundamental Market

It might be thought that our fundamental market behaves exactly as the EMH predicts the stock market behaves, so there is nothing new here. However, when we discussed the EMH in Section 7, we stressed the desirability of extending it to include nonzero transactions and calculation costs and heterogeneous rational agents, and we would want the fundamental market that is part of the GTSM to have these features as well.

Transactions costs and calculation costs mean there is a 'band' around the fundamental value within which the share price can fluctuate independently of changes in fundamentals. Heterogeneous rational agents mean that the demand curve for shares almost certainly slopes downwards, so that, for example, an exogenous increase in the supply of shares would depress share prices. However, there are also complications due to the possibility that the economy might move from the fundamental state to a bull or bubble state.

Firstly, we might mention what we might describe as 'bull market risk'. Suppose a firm's share price rises above its fundamental value by an amount that exceeds the relevant transactions and calculation costs, so that one would expect rational agents to sell. But a

rational trader may not sell the share if he thinks a bull market may be about to start – i.e. if he suspects there may be further increases in the share price above its fundamental value. Perhaps this may raise the ‘ceiling’ of the band without affecting the floor.

Secondly, agents may expect that, with some probability, the share price may become overvalued in the future because it might enter a bull or bear market. It is clear that such a possibility raises the price a rational agent should be willing to pay for the share above its current fundamental value. So share prices in the fundamental market might incorporate what we might describe as an ‘overvaluation premium’ which reflects this possibility. Notice that this differs from ‘bull market risk’, which arises when the share price rises; the overvaluation premium might always exist, raising share prices whether the market is rising or falling. It might often be fairly small or non-existent, but at times might become quite large, although this is conjecture.

We might also ask whether heterogeneity of rational expectations is compatible with it being impossible to systematically beat the market. This is undoubtedly something that requires further work. We would point out that the fact that expectations may be different does not mean that one way of forming expectations is always better than another – one may be better than another in some circumstances but not in others. Also, we would expect that if there is an obvious way of beating the market this should be incorporated into all rational agents’ expectations. If one agent has expectations that seem to do consistently better than other agents, one would expect the agents whose expectations perform worse would try to improve them, perhaps by copying or investigating the expectations of those whose expectations seem to succeed. So heterogeneous rational expectations, which might be expected to change over time as agents learn, should be compatible with the result that it

is impossible to beat the market systematically, but on the other hand would not be incompatible with there being an investor who, perhaps because of superior skill or information processing ability, would be able to make excess returns. But such individuals might be fairly rare.

12. How does the GTSM Explain the Anomalies?

That share prices move in response to news is certainly compatible with the GTSM, although we need to understand more about how and why news causes share prices to move in both bull and bear markets. And the GTSM is also compatible with the idea that it is impossible to (systematically) beat the market. It is also clearly compatible with the views of stock market professionals who believe in bull and bear markets. Here is how the theory is capable, at least in principle, of explaining the other anomalies:

Excess Valuations. Clearly the GTSM is compatible with excess valuations of shares. In both the bull and bear market regimes, prices exceed their fundamental values.

Excess Volatility. According to the GTSM, share prices are sometimes equal to their fundamental values and sometimes above, but never below. This does not necessarily imply excess volatility, but it is easy to give a condition under which we would see excess volatility. Since the share price at time t (P_t) must, by definition, equal the sum of its fundamental value (F_t) and its bubble or excess value (B_t), then the variance of the share price is given by

$$\text{var}(P_t) = \text{var}(F_t) + \text{var}(B_t) + 2\text{cov}(F_t, B_t). \quad (5)$$

So there will be excess volatility unless the covariance between the fundamental and bubble value is negative and exceeds, in absolute value, half the variance of the bubble term. One would expect that this condition for excess volatility is usually satisfied. In particular, one

would expect the covariance between the fundamental value and bubble value of a share to be positive if bubbles tend to arise when there is an increase in fundamental values. It would seem then that the GTSM is capable of explaining excess volatility. Whether it can explain the amount of excess volatility exhibited is another matter and one would want to see this demonstrated, perhaps in a properly calibrated model of the GTSM.

Bubbles. Clearly, the GTSM is compatible with the existence of bubbles, but of course it needs to be explained how and why bubbles emerge and how and why they end. (More generally, there needs to be an account of the circumstances under which one type of market might change to another type.)

Crashes. These can occur under the GTSM for three reasons: (i) a large decline in fundamentals in a fundamental market; (ii) the ending of a bull market (collapse of a bubble); (iii) a fall in share prices because of a decision of a number of noise traders to sell in a bear market. So it is not at all implausible that we should occasionally see large downward movements in share prices without there being a large deterioration of fundamentals.

Calendar Anomalies. These can perhaps be explained as movements within the band in a fundamental market due to such factors as sentiment. The anomalies seem to be too small to generate excess returns after transactions costs, etc., are taken into account. Also, there is no reason to suppose that such effects will not operate in bull and bear markets.

Momentum Effects. Clearly there can be momentum effects in the upward direction in a bull market and in the downward direction in a bear market.

Mean Reversion. If shares are sometimes at their fundamental levels, but sometimes experience bubbles which then burst and become bear markets before reverting over time to their fundamental levels, then we would expect to see mean reversion.

Downward-Sloping Demand Curves for Stocks. Shleifer's evidence that share prices move in response to an exogenous change in demand can readily be explained by heterogeneous expectations in the fundamental state; it might be conjectured that exogenous changes in supply or demand might have similar effects in bull and bear markets, although this is something that does require confirmation.

Overreaction and Underreaction. A share price may overreact to an improvement in fundamentals if the purchase of the shares by rational agents leads to irrational agents extrapolating the price increase and bidding the price above their fundamental value. A share price may underreact to a negative shock to fundamentals if purchases by irrational agents prevent the share price from falling all the way to its new fundamental level. Generally, a whole number of patterns of adjustment are possible in the GTSM.

Stop-Loss Rules. Perhaps minor price downturns can occur in a bubble without signalling that the bubble is over (in which case it would not be rational to sell), whereas a larger downturn may signal that the bubble is over (and hence it may be rational to sell). Sometimes a stop-loss rule of 20% is recommended. It might be that practitioners have learned that a price decline of this magnitude generally signals the start of a bear market, so it might be rational to sell after such a price decline.

Trading Volume. A model with heterogeneous agents (some irrational) clearly offers much more promise to explain the observed volume of trading in the stock market. If a shock changes the valuations that different agents put on a share price, it is easy to see that this may give rise to trading, and this effect might be reinforced if some agents are irrational and interpret the implications of the shock very differently from the way in which rational agents do.

13. Methodological Implications

The assumption that all agents are rational is often made in macroeconomics and in many other areas of economics. In response to abundant evidence that many agents cannot plausibly be described as rational in many of the decisions they make, advocates of rational expectations have often argued that their results go through provided that ‘enough’ agents are rational. However, while models where just a fraction of agents are rational may under certain circumstances generate the same results as models where all agents are rational, under some other conditions they do not. Instead of assuming that all agents are rational in models, perhaps we should assume that only some are, and equilibrium is where rational agents cannot make themselves better off.

However, the concept of rationality is quite complicated. Rationality seems a reasonable assumption for many of our decisions. Most of us would (say) choose a deposit account paying 3% to an identical deposit account paying 2%. When the decisions become more complicated, though, the assumption of rationality becomes more dubious. We might think of rational decision making as consisting of weighing the evidence and making a decision that tends to bring about whatever goal we might wish to achieve. If rationality is envisaged in such a way, there is no difficulty in supposing that rational agents may differ in their expectations. And although some basic rationality seems reasonable, the very demanding form of rationality that seems to be required for some results in economics (such as the common knowledge assumption) seems quite unreasonable. Nevertheless, even if some assumptions of rationality seem unreasonable, it is not often clear how to relax them. Appendix II discusses these issues at greater length.

The upshot is that we do need to seriously consider models where some agents have rational, yet heterogeneous expectations, but others have irrational expectations. Modelling these expectations and how they might change over time is clearly no trivial task. But also, working out how overall equilibrium is determined with such heterogeneity of agents is a major challenge as well. What our discussion has suggested is that the presence of irrational agents and heterogeneous rational expectations may make little difference in some circumstances, but may make a big difference in others. Building appropriate models to explore this issue is an important item on the research agenda.

A further point is that once one allows for heterogeneity of expectations, the decisions of rational agents become much more complicated. It is no longer sufficient for them to just attempt to work out a share's fundamental value (difficult though that may be) when deciding whether to buy a share. They also need to try to predict what other agents are likely to do, and such prediction requires them to try to predict what these agents predict other agents will do, and so forth. Clearly, building such models may be extremely difficult indeed. Nevertheless, we may often need to go beyond the assumption of a homogenous representative agent. The current paper suggests that models that relax this assumption in some particular way – say by admitting a mixture of rational and irrational agents, or by assuming heterogeneous rational agents, may possess considerable additional explanatory power.

14. Empirical Evidence

In this section we mention a few empirical studies the results of which are particularly relevant for the approach we are developing.

Guidolin and Timmermann (2005) present ‘empirical evidence of regimes with very different risk and return characteristics for UK stocks and bonds. Our results suggest that a three-state specification with a transitory high-volatility regime with negative mean returns, a highly persistent, ‘normal’ state with mean returns and volatility levels close to historical averages and a persistent high-return ‘bull’ state capture important features of UK stock and bond returns.’ (*op. cit.*, p. 141.)

This seems to fit the GTSM exactly! However the paper is entirely atheoretical – the authors make no effort at all to explain their results in terms of any theory.

In a recent paper Ahn *et al* (2014) present evidence that the shares of Motors Liquidation Company (the company that used to be known as General Motors) were overpriced. They develop a theoretical asset-pricing model that includes both rational speculators and uninformed investors, which has a state where all rational agents exit and the share price is inflated. This seems to correspond exactly with our bear market state.

There are several papers which study the effects of dispersion in investors’ beliefs for stock prices. One example is Diether, Malloy and Scherbina (2002).

15. Implications for Investor Behaviour

The GTSM may not have radically different implications for investor behaviour from the EMH, but it does have some. Certainly in the fundamental state, a ‘buy and hold’ strategy seems optimal, with periodic ‘rebalancing’ of the portfolio in order to preserve adequate diversification (as is also an implication of the EMH). However, at any one point in time, some shares may be overvalued and hence should not be bought; it may take some research in order to identify them.

If the market is in a bubble state, then it is quite rational to participate. However, it is not clear whether rational investors should be able to predict the end of a bubble. If they can, then it would seem to be a clear implication that one should sell in those circumstances. In a bear state, presumably the implication is that one should not buy shares; indeed, one should sell the shares one has; however it is perhaps implausible for all shares to be overvalued. Instead, it may be advisable to avoid overvalued shares and purchase shares that are valued at their fundamental value.

16. Further Issues

There are many issues to which the GTSM might be applied and on which it may give useful insights.

One question is whether it applies to individual stocks or to the stock market as a whole. It should be quite clear that different shares can be in different states. For example, there is nothing in the approach which precludes some shares being priced at their fundamental values and others being above their fundamental levels, being either in a bear market or bubble. Rational agents might hence hold both fundamentally priced shares and those in a bubble state. Irrational agents would be the sole holders of shares that are in a bear market state, but they might hold both fundamentally priced shares and bull market shares as well.

Another interesting question is whether the approach is applicable to other asset markets such as housing. It is, of course, quite common to talk of 'housing price bubbles' and so forth. However, heavy transactions costs in buying and selling houses means that a basic arbitrage equation such as equation (1) is unlikely to hold for housing, even approximately – instead, one would get a pair of inequality constraints.

The approach might have implications for the international correlation of stock markets. In a fundamental market, shares prices might be correlated internationally because of common shocks – i.e. changes in fundamentals that affect many countries. If shares can be in a bull or bear market, then other possibilities emerge. In a bull market, there might be international transmission of sentiment, so share prices might be more highly correlated than in a fundamental market. However, in a bear market, it is not clear that shocks to fundamentals will change share prices, so prices might be much less correlated internationally.⁵

Sometimes a tax on trading such as a Tobin tax has been advocated as a way of reducing stock market volatility (e.g. see the discussion in Shiller, 2000, pp. 225 – 8). It is easy to see that such a tax should raise volatility in a fundamental market, as the greater the costs of buying or selling a share, the more an undervalued share will have to fall below its fundamental value before rational speculators have an incentive to buy it. But in a bull market it may be different. A transactions tax may deter speculators from buying a share and hence might prevent a bubble from occurring, or at least might reduce its magnitude. So the effect of such a tax on share price volatility might well depend on the regime the stock market is in. Clearly, this is something that needs to be explored in a formal model.

17. Summary of Main Points

1. The stock market may be in one of three states: a fundamental state in which share prices (approximately) equal their fundamental values, a bull market in which prices exceed their fundamental value but the efficient market condition (1) holds, and a bear market where equation (1) does not hold and shares are overvalued.

⁵ Quinn and Voth (2008) provide some evidence on the international correlation of stock market returns.

2. In the fundamental state shares may not exactly equal their fundamental values because of transactions and calculation costs. Shares may also be somewhat above their fundamental values because of the expectation (by rational agents) that the shares might enter a bull market or bear market state.
3. Bubbles can occur. In developing models of bubbles the extreme rationality assumptions that are often made in models of bubbles need to be relaxed; appropriate models of bubbles should hence be developed assuming a mixture of rational and irrational agents.
4. In a bear market, the marginal investor is irrational. It could be that all shares are held by irrational agents, with rational agents being unable to exploit the overvaluation because of short-selling constraints. But optimistic rational agents could hold shares in a bear market, provided that the marginal investor is irrational.
5. It is also important to consider the implications of assuming heterogeneous rational agents in the fundamental market.

What is new about the GTSM? Although most of the 'ingredients' of the GTSM are not new (e.g. the point about short-selling constraints giving rise to overvaluation goes back to Miller, 1977), we would claim the way in which the ingredients are combined to produce the overall theory is novel. We would point out that in the theory shares can be overvalued for two very different reasons. The first is that they are overvalued because they are expected to rise still further in price (i.e. there is a bubble). The second is that they are overvalued because they are held exclusively by irrational agents and rational agents cannot exploit the overvaluation because of short-selling constraints. The two states in which share prices are overvalued can be expected to behave in very different ways. In the first, share prices might be expected to rise on average over time and in the second they might be

expected to fall. It is not clear that this possibility (that overvalued shares may behave in very different ways) has been recognised in the literature. We have also emphasised heterogeneity of investors' expectations. Discussion of such heterogeneity is not of course absent from the literature. One important paper is De Long *et al.* (1990) who consider the interaction of irrational noise traders and rational arbitrageurs. Hong and Stein (2007) also survey much of the literature on investor heterogeneity. We would stress that this heterogeneity, which certainly exists, may affect the behaviour of the stock market in different ways in different circumstances. Again, it is not clear that this point has been recognised in the literature.

18. Conclusions

The paper has argued that the stock market can best be understood as being comprised of a mixture of rational and irrational agents. There are several possible outcomes for the way the stock market behaves. The behaviour of the stock market may be determined entirely by the behaviour of rational agents; alternatively, there may be circumstances where share prices are determined by the decisions of irrational agents and rational agents cannot exploit any overvaluation this may give rise to. It has also been argued that there may be heterogeneous rational expectations; such an assumption does not contradict any basic principle of economics and may play a role in explaining the behaviour of the stock market.

So the GTSM combines the insights of Fama, who argues for the efficiency of the stock market (it may be approximately efficient for much of the time under the GTSM) with the insights of Shiller (it can explain excess volatility, for example). There is plenty of room for behavioural economics to explain the behaviour of irrational agents, and the approach

can potentially integrate the insights of behavioural economics into an overall theory of the stock market. It should be able to construct a model which explains the reaction of stock prices to (say) an expansion by the central bank of its quantitative easing programme, something which behavioural economics by itself does not seem able to explain, so hence providing a genuine alternative to the EMH, and meeting Fama's point quoted above. The theory should be able to explain the anomalies which the EMH finds extremely difficult to explain (excess volatility, overvaluation, bubbles, momentum effects, overreaction and underreaction), but is also compatible with it being impossible to systematically beat the market and with the stock market reacting to news.

There is much work which needs to be done to develop such a theory, but we would claim that it does contain the essence of a satisfactory theory of the stock market.

APPENDIX I

An Example of Overvaluation

Consider a firm which has just announced the following results:

- (1) In the year for which it is reporting results, it had revenues of £23.3m and a loss of £31.4m.
- (2) In the previous year it had revenues of £14.84m and a loss of £5.17m.
- (3) It has net cash of £96m.

Also, it is regarded as a 'world leader' in a rapidly growing sector of the economy.

Question: what would be a plausible estimate of the market value of the company?

According to the Efficient Markets Hypothesis (EMH), share prices are determined as in equation (3), which we repeat here for convenience:

$$P_0 = \sum_{i=1}^{i=\infty} \left(\frac{E[D_i]}{(1+r)^i} \right) \equiv F_0. \quad (3)$$

Let us try to estimate a plausible upper bound on the market value of the company. Consider the following scenario:

1. In the next three years, the firm's revenues grow to £100m., using its cash reserves which it exhausts at the end of the three years (so revenues double every 17 months or so in this period).
2. In the following four years, its revenues grow from £100m to £400m. (i.e., they double every two years.) Assume it finances its expansion by internally generated funds.
3. Thereafter, the firm's revenues grow at 5% per year indefinitely, and it pays a dividend equal to 10% of its revenues (so its first dividend is £40m.)

Then it is not too difficult to calculate that the firm's market value in 7 years' time, using a discount rate of 10%, will be £800m. This means its current market value will be about £400m.

Is this credible?

The above calculations are based on some **extremely** favourable assumptions about the firm's behaviour. For example, it is assumed that its average rate of growth of revenues will be about 35% over the next ten years. According to Chan et al., 2003, only 2% of firms have annual sales growth rates in excess of 22% over a ten-year period, and these figures are biased upward by survivorship bias. Also, assuming that the firm can pay dividends equal to 10% of revenues indefinitely after ten years seems very strong indeed.

So it would seem that £400m is a considerable overestimate of what the firm's market value should be; indeed it would seem to be a considerable overestimate of the

maximum plausible market value of the company, assuming that its price equals its fundamental value.

However, perhaps it can now be revealed the firm in question is one called Baltimore Technologies, and the figures are those for March 2000. At the time its market value was £4.48bn, that is about **eleven times** a value which was argued to far exceed its fundamental value! It seems very difficult to avoid the conclusion that Baltimore Technologies was (massively) overvalued.

APPENDIX II

On the Rationality and Heterogeneity of Expectations

We would make a number of contentions about expectations here.

- (1) There is often a huge amount of heterogeneity in expectations.
- (2) For some decisions, an assumption that virtually all agents making the decisions have rational expectations is defensible. For many other decisions, it is reasonable to suppose that some agents have rational expectations and other do not. Furthermore, there are some decisions for which an assumption of rational expectations is much less plausible for any agent.
- (3) It is plausible that those expectations, which may plausibly be regarded as 'rational', can differ between agents, sometimes massively.
- (4) As far as investors in the stock market is concerned, many investors make buying and selling decisions that in no reasonable way can be considered rational.

To make a compelling case for these contentions would take a long time. The rest of this Appendix contains some discussion which is hopefully relevant although it may fall short of making a persuasive case.

On contention (1), for evidence on heterogeneity of shareholder expectations, see Bagwell (1991) and Booth (1993), footnote 7.

As far as contention (2) is concerned, many decisions we make can plausibly be described as rational. For example, virtually everyone would pick up a £20 bill on the pavement (sometimes thought of as the archetypal rational decision). There are a number of well-established results in economics which hold because some agent can make a clear-cut unambiguous gain from acting in a particular way. One example is covered interest parity, which holds because some agents perceive and act upon opportunities to make riskless arbitrage profits. But for many other decisions, rationality seems much less plausible. Behavioural economists and psychologists have documented numerous biases and distortions in much of our decision making; Kahneman (2011) is the obvious reference.

However, there are some decisions where rationality is much less plausible for any agent, or only is a reasonable assumption for a handful of specialist and exceptionally well-informed individuals. Examples might include the level of rationality that each agent needs to possess if Ricardian equivalence is to hold, and the information some speculators need to possess and the calculations they need to do to rule out divergent dynamic paths in some optimizing macroeconomic models.

Contention (3), that there may be heterogeneity in rational expectations, may seem a strange idea, and has scarcely featured in the literature, but it may be justified by supposing that each agent may have idiosyncratic information. Alternatively, it might be that it is costly to gather information, and agents gather different amounts of information, and hence may

have different expectations, because their costs of gathering information differ.⁶ At a casual level, we might note that many stockbrokers have widely differing recommendations on the same shares and it would be an uncontroversial statement that many well-informed people have radically divergent views on the same issues. To take the example that started the paper, Fama and Shiller are both outstanding academics who have spent their careers studying the stock market – if we do not consider them as ‘rational’ then who is? Yet they have very different views on how the stock market operates. This is surely one example of heterogeneous rational expectations.

As far as contention (4) is concerned there is plenty of evidence that many traders in the stock market can by no means plausibly be described as rational. Here are some possible candidates for irrational or noise traders:

(i) Managers of Index Funds. A considerable proportion of shares are held in index funds. Managers of index funds construct their portfolios so that the proportions of shares in these portfolios are the same as the weights of the shares in the index they are tracking, so if a particular share is given a weight of say 5% in the index, the fund will hold 5% of its portfolio in that share. Index funds are quite popular, and many funds that are not explicit index funds are believed to be ‘closet’ index trackers. Investing in tracker funds has an obvious attraction for believers in the EMH. If one cannot systematically beat the market, there is no point in wasting resources researching various shares - one may as well buy shares in an index fund, which hopefully can minimise transactions costs. However, this means that managers of index funds buy and sell shares for reasons completely unrelated to whether they are over- or under-valued by fundamental criteria; the sole consideration is replicating the index. So if

⁶ See Brunnermeier and Parker (2005) for one model in which agents may optimally have heterogeneous expectations.

a new company enters the index, the index fund has to buy the shares of that company even if its shares are overvalued.

(ii) Institutional Investors Motivated by Relative Performance Criteria. Institutional investors are very important in the stock market, and it is commonly believed that they are motivated by relative performance criteria – i.e. that what matters is not how well or badly they perform in an absolute sense but how well or badly they perform relative to other institutional investors. This means that they shift their portfolios towards stocks that comprise their benchmark index. (See Basak and Pavlova, 2013.)

(iii) Small Individual Traders. De Bondt (1998) presents survey evidence according to which many individual investors behave very differently from the way in which standard models predict.

(iv) Technical Analysts. Such analysts try to predict stock price changes on the basis of past patterns in stock prices. An elaborate vocabulary has grown up around technical analysis; terms such as ‘resistance level’, ‘support level’ and ‘head and shoulders pattern’ are commonly used. Typically, economists have been quite sceptical about technical analysis, sometimes asserting that it has the scientific status of astrology. Yet technical analysis is widely used by financial market professionals; according to one survey: ‘At a forecasting horizon of weeks, technical analysis is the most important form of analysis.’ (Menkhoff, 2010, p. 2573.)

These may be the main categories of noise or irrational traders; there are undoubtedly others. We might ask whether traders who only review their portfolios periodically should be considered irrational. Ascertaining fundamental values may be quite difficult and costly, so it may be rational for many investors to gather information about

their portfolios and whether their shares are over or undervalued periodically. But the archetypal rational investor needs to be constantly updating his or her information and be ready to trade when they perceive a mis-valuation. It is difficult to believe that many investors come anywhere close to doing this.

This Appendix is intended to make the case that for many decisions, including stock market investment decisions, there is a huge amount of heterogeneity. We may make a distinction for analytical purposes between rational and irrational traders, which may shed some insights, but even this distinction is simplistic. (We would expect many noise traders to act rationally in many situations, such as picking up £20 notes on the pavement.)

Characterising the heterogeneity of agents, and developing models of the stock market which incorporate this heterogeneity, is clearly no easy task. It needs to be an important part of the research agenda over the next few years.

REFERENCES

- Abreu, Dilip and Marcus K Brunnermeier, 'Bubbles and Crashes', *Econometrica*, 2003, 71 (1), pp. 173 – 204.
- Ahn, Tom, Jeremy Sandford and Paul Shea, 'A Note on Bubbles, Worthless Assets and the Curious Case of General Motors', *Macroeconomic Dynamics*, 2014, 18 (1), pp. 244 – 54.
- Bagwell, Laurie S., 'Shareholder Heterogeneity: Evidence and Implications', *American Economic Review*, 1991, 81 (2), pp. 218 – 221.
- Basak, Suleyman and Anna Pavlova, 'Asset Prices and Institutional Investors', *American Economic Review*, August 2013, 103 (5), pp. 1728 – 58.

- Blanchard, Olivier and Mark Watson, 'Bubbles, Rational Expectations and Financial Markets', in Paul Wachtel (ed.) *Crises in the Economic and Financial Structure*, Lexington, MA, Lexington Books, 1982.
- Booth, Richard A., 'The Efficient Market, Portfolio Theory, and the Downward Sloping Demand Hypothesis', *New York University Law Review*, 1993, 68 (5), pp. 1187 – 1212.
- Brunnermeier, Marcus K. and Stefan Nagel, 'Hedge Funds and the Technology Bubble', *Journal of Finance*, October 2004, 59 (5), pp. 2013 - 2040.
- Brunnermeier, Markus K. and Jonathan A. Parker, 'Optimal Expectations', *American Economic Review*, 2005, 95 (4), pp. 1092 – 1118.
- Campbell, John Y. 'Empirical asset Pricing: Eugene Fama, Lars Peter Hansen and Robert Shiller', *Scandinavian Journal of Economics*, 2014, 116 (3), pp. 593 - 634.
- Chan, Louis K. C., Jason Karceski and Josef Lakonishke, 'The Level and Persistence of Growth Rates', *Journal of Finance*, 2003, 58 (2), pp. 643 – 684.
- De Bondt, Werner, 'A Portrait of the Individual Investor', *European Economic Review*, 1998, 42(3-5), pp. 831–844.
- De Long, J. Bradford, Andrei Shleifer, Lawrence Summers and Robert Waldmann, 'Noise Trader Risk in Financial Markets', *Journal of Political Economy*, 1990, 98 (4), pp. 703 – 728.
- Diether, Karl, Christopher Malloy and Anna Scherbina, 'Differences of Opinion and the Cross-Section of Stock Returns', *Journal of Finance*, 2002, 57 (5), pp. 2113 – 41.
- Fama, Eugene, 'Two Pillars of Asset Pricing', *American Economic Review*, 2014, 104(6), pp. 1467 – 1485.
- Fender, John, 'Stop-losses, Do they Work?' *Investors Chronicle*, 13 – 19 July 2001, pp. 23 – 25.

- Garber, Peter, 'Famous First Bubbles', *Journal of Economic Perspectives*, 1990, 4 (2), pp. 35 - 54.
- Grossman, Sanford J. and Joseph Stiglitz, 'On the Impossibility of Informationally Efficient Markets', *American Economic Review*, 1980, 70 (3), pp. 393 – 408.
- Guidolin, Massimo and Allan Timmermann, 'Economic Implications of Bull and Bear Regimes in UK Stock and Bond Returns', *Economic Journal*, 2005, 115 (January), pp. 111 – 43.
- Haug, Mark and Mark Hirschey, 'The January Effect', *Financial Analysts Journal*, 2006, 62(5), pp. 78 – 88.
- Hong, Harrison and Jeremy Stein, 'Differences of Opinion, Short-Sales Constraints, and Market Crashes', *Review of Financial Studies*, 2003, 16(2), pp. 487 – 525.
- Hong, Harrison and Jeremy Stein, 'Disagreement and the Stock Market', *Journal of Economic Perspectives*, 2007, 21 (2), pp. 109 – 128.
- Jegadeesh, Narasimhan and Sheridan Titman, 'Profitability of Momentum Strategies: An Evaluation of Alternative Explanations', *Journal of Finance*, 2001, 56 (2), pp. 699 – 720.
- Jensen, Michael, 'Some Anomalous Evidence Regarding Market Efficiency', *Journal of Financial Economics*, 1978, 6 (2 – 3), pp. 95 – 101.
- Jones, Charles M. and Owen A. Lamont, 'Short-Sale Constraints and Stock Returns', *Journal of Financial Economics*, 2002, 66 (2 – 3), pp. 207 – 39.
- Kahneman, Daniel, *Thinking, Fast and Slow*, 2011, London, Allen Lane.
- Kuhn, Thomas S. *The Structure of Scientific Revolutions*, 1970, Chicago, University of Chicago Press, Second Edition.

- Malkiel, Burton, 'The Efficient Markets Hypothesis and Its Critics', *Journal of Economic Perspectives*, 2003, 17(1), pp. 59 – 82.
- Menkhoff, Lukas, 'The Use of Technical Analysis by Fund Managers: International Evidence', *Journal of Banking & Finance*, 2010, 34, pp. 2573 – 2586.
- Miller, Edward M., 'Risk, Uncertainty and Divergence of Opinion', *Journal of Finance*, 1977, 32(4), pp. 1151 – 1168.
- Ofek, Eli and Matthew Richardson, 'The Valuation and Market Rationality of Internet Stock Prices', *Oxford Review of Economic Policy*, 2002, 18 (3), pp. 265 – 287.
- Poterba, James and Larry Summers, 'Mean Reversion in Stock Prices: Theory and Evidence', *Journal of Financial Economics*, 1988, 22 (1), pp. 27 – 59.
- Quinn, Dennis P. and Hans-Joachim Voth, 'A Century of Global Equity Market Correlations', *American Economic Review: Papers & Proceedings*, 2008, 98 (2), pp. 535 – 540.
- Santos, Manuel and Michael Woodford, 'Rational Asset Pricing Bubbles', *Econometrica*, 1997, 65 (1), pp. 19 – 57.
- Scheinkman, José and Xiong, Wei, 'Overconfidence and Speculative Bubbles', *Journal of Political Economy*, 2003, 111(6), pp. 1183 - 1219.
- Scherbina, Anna and Bernd Schlusche, 'Asset Price Bubbles: A Survey', *Quantitative Finance*, 2014, 14(4), pp. 589 – 604.
- Shiller, Robert J. 'Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?' *American Economic Review*, 1981, 71 (3), pp. 421 – 35.
- Shiller, Robert J., *Irrational Exuberance*, 2000, Princeton, Princeton University Press, First Edition.

Shiller, Robert J., 'Speculative Asset Prices', *American Economic Review*, 2014, 104(6), pp. 1486 – 1517.

Shleifer, Andrei, 'Do Demand Curves for Stocks Slope Downwards?' *Journal of Finance*, 1986, 41 (3), pp. 579 – 590.

Shleifer, Andrei, *Inefficient Markets: an Introduction to Behavioural Finance*, 2000, Oxford, Oxford University Press.

Siegel, Jeremy, *Stocks for the Long Run*, 2002, New York, McGraw-Hill, Third Edition.

Temin, Peter and Hans-Joachim Voth, 'Riding the South Sea Bubble', *American Economic Review*, 2004, 94(5), pp. 1654 – 68.

Tirole, Jean, 'On the Possibility of Speculation under Rational Expectations', *Econometrica*, 1982, 50 (5), pp. 1163 – 81.

Tirole, Jean, 'Asset Bubbles and Overlapping Generations', *Econometrica*, 1985, 53(5), pp. 1071 – 1100.

White, Eugene N., 'The Stock Market Boom and Crash of 1929 Revisited', *Journal of Economic Perspectives*, 1990, 4 (2), pp. 67 - 83.