A Survey of Behavioral Finance

Nicholas Barberis and Richard Thaler

August 2001

Abstract

Behavioral finance argues that some financial phenomena can plausibly be understood using models in which some agents are not fully rational. The field has two building blocks: *limits to arbitrage*, which argues that it can be difficult for rational traders to undo the dislocations caused by less rational traders; and *psychology*, which catalogues the kinds of deviations from full rationality we might expect to see. We discuss these two topics, and then present a number of behavioral finance applications: to the aggregate stock market, to the cross-section of average returns, to individual trading behavior, and to corporate finance. We close by assessing progress in the field and speculating about its future course.

Contents

1 Introduction

The traditional finance paradigm, which underlies many of the other articles in this handbook, seeks to understand financial markets using models in which agents are rational. In finance, "rationality" means two things. First, agents' beliefs are correct: the subjective distribution they use to forecast future realizations of unknown variables is indeed the distribution that those realizations are drawn from. Second, given their beliefs, agents make choices that are normatively acceptable, in the sense that they are consistent with Savage's notion of Subjective Expected Utility (SEU).

This traditional framework is appealingly simple, and it would be very satisfying if its predictions were confirmed in the data. Unfortunately, after years of effort, it has become clear that basic facts about the aggregate stock market, the cross-section of average returns, and individual trading behavior are not easily understood in this framework.

Behavioral finance is a new approach to financial markets that has emerged, at least in part, in response to the difficulties faced by the traditional paradigm. In broad terms, it argues that some financial phenomena can be better understood using models in which some agents are *not* fully rational. More specifically, it analyzes what happens when we relax one, or both, of the two tenets that underlie the finance view of rationality. In some behavioral finance models, agents hold beliefs that are not completely correct, most commonly because of a failure to apply Bayes' law properly. In other models, agents hold correct beliefs but make choices that are normatively questionable, in that they are incompatible with SEU.

This review essay evaluates recent work in this rapidly growing field. In Section 2, we consider the classic objection to behavioral finance, namely that even if some agents in the economy are irrational, rational agents will prevent them from influencing security prices for very long, through a process known as arbitrage. One of the biggest successes of behavioral finance is a series of theoretical papers showing that in an economy where rational and irrational traders interact, irrationality *can* have a substantial and long-lived impact on prices. These papers, known as the literature on "limits to arbitrage," form one of the two buildings blocks of behavioral finance.

In order to make sharp predictions, behavioral models often need to specify the form of agents' irrationality. How exactly do people misapply Bayes law or deviate from SEU? For guidance on this, behavioral economists typically turn to the extensive experimental evidence compiled by cognitive psychologists on the biases that creep in when people form *beliefs*, and on people's *preferences*, or on how they make decisions, given their beliefs. Psychology is therefore the second building block of behavioral finance, and we review the psychology most relevant for financial economists in Section 3.

In Sections 4-8, we consider specific applications of behavioral finance: to understanding the aggregate stock market, the cross-section of average returns, and the pricing of closed-end funds in Sections 4, 5 and 6 respectively; to understanding how particular groups of investors choose their portfolios and trade over time in Section 7; and to understanding the security issuance, capital structure, and dividend policies of firms in Section 8. Section 9 takes stock and suggests directions for future research.[1]

2 Limits to Arbitrage

2.1 Market Efficiency

In the traditional finance paradigm where agents are rational, security prices equal "fundamental value." This is the discounted sum of expected future cashflows, where the expectation is taken over the correct distribution and where the discount rate is consistent with a normatively acceptable preference specification. The hypothesis that actual prices equal fundamental value is known as the Efficient Markets Hypothesis (EMH). Put simply, under this hypothesis, "prices are right," in that they are set by rational agents. In an efficient market, there is "no free lunch": no investment strategy can earn excess risk-adjusted returns, or average returns greater than are warranted for its risk.

Behavioral finance argues that some features of asset prices are most plausibly interpreted as deviations from fundamental value, and that these deviations are brought about by the presence of irrational traders in the economy. A long-standing objection to this view that goes back to Friedman (1953) is that rational traders will quickly undo any dislocations caused by irrational traders. To illustrate the argument, suppose that the fundamental value of a share of Ford is \$20. Imagine that a group of irrational traders becomes excessively pessimistic about Ford's future prospects and through its selling, pushes the price to \$15. Defenders of the EMH argue that rational traders, sensing an attractive opportunity, will buy the security at its bargain price and at the same time, hedge their bet by shorting a "substitute" security, or another security that has similar cashflows to Ford in future states of the world. The buying pressure on Ford shares will then bring their price back to fundamental value.

Friedman's line of argument is initially compelling, but it has not survived careful theoretical scrutiny. In essence, it is based on two assertions. First, as soon as there is a deviation from fundamental value – more simply, a mispricing – an attractive investment opportunity is created. Second, rational traders will immediately snap up the opportunity, thereby correcting the mispricing. Behavioral finance does not take issue with the second step in this argument: when attractive investment opportunities come to light, there is lit-

[1] We draw readers' attention to two other recent surveys of behavioral finance. Both are excellent. Shleifer (2000) provides a particularly detailed discussion of the theoretical and empirical work on limits to arbitrage, which we summarize in Section 2. Hirshleifer's (2001) survey is closer to ours in terms of material covered, although we devote less space to asset pricing applications, and more to corporate finance and individual trading applications. We also organize the material somewhat differently.

tle doubt that they are quickly exploited. Rather, it disputes the first step. The argument, which we elaborate on in Sections 2.2 and 2.3., is that even when an asset is wildly mispriced, strategies designed to correct the mispricing can be very risky, rendering them unattractive. As a result, the mispricing can remain unchallenged.

It is interesting to think about common finance terminology in this light. While irrational traders are often known as "noise traders," rational traders are typically referred to as "arbitrageurs." Strictly speaking, an arbitrage is an investment strategy that offers riskless profits at no cost. Presumably, the rational traders in Friedman's fable became known as arbitrageurs because of the belief that a mispriced asset immediately creates an opportunity for riskless profits. Behavioral finance argues that this is *not* true: the strategies that Friedman would have his rational traders adopt are not necessarily arbitrages; quite often, they are very risky strategies.

An immediate corollary of this line of thinking is that "prices are right" and "there is no free lunch" are *not* equivalent statements. While both are true in an efficient market, "no free lunch" can also be true in an inefficient market: just because prices are away from fundamental value does not necessarily mean that there are any excess risk-adjusted returns for the taking. In other words,

$$\boxed{\text{“prices are right”} \Rightarrow \text{“no free lunch”}}$$

but

$$\boxed{\text{“no free lunch”} \not\Rightarrow \text{“prices are right”}.}$$

This distinction has important implications for the current debate on market efficiency. First, many researchers (see Ross, 2001) point to the inability of professional money managers to beat the market as strong evidence of market efficiency. However, underlying this argument is the assumption that "no free lunch" implies "prices are right." If, as we argue in Sections 2.2 and 2.3, this link is broken, the performance of money managers is not an accurate guide to the efficiency of markets.

Second, while some researchers accept that there is a distinction between "prices are right" and "there is no free lunch," they believe that the debate should be more about the latter statement than about the former (Rubinstein, 2000). We disagree with this emphasis. Whether or not a market contains free lunches, our concern as economists should be about whether prices are right: only then can we be sure that capital is being correctly allocated to the most promising investment opportunities.

2.2 Theory

In the previous section, we emphasized the idea that when a mispricing occurs, strategies designed to correct it can be very risky, allowing the mispricing to survive. Here we discuss four sources of risk that have been identified in the literature. In our discussion, we return to the example of Ford, whose fundamental value is $20, but which has been pushed down to $15 by pessimistic noise traders.

Fundamental Risk

The most obvious risk that an arbitrageur who buys Ford's stock at $15 faces is that a piece of bad news about Ford's fundamental value causes the stock to fall further, leading to losses. Of course, arbitrageurs are well aware of this risk, which is why they short a substitute security such as General Motors at the same time that they buy Ford. The problem is that substitute securities are rarely perfect, and often highly imperfect, making it impossible to remove all the fundamental risk. Shorting General Motors protects the arbitrageur somewhat from news about the car industry as a whole, but still vulnerable to news that is specific to Ford – news about defective tires, say.

Noise Trader Risk

Noise trader risk, an idea introduced by De Long et. al. (1990a) and studied further by Shleifer and Vishny (1997), is the risk that the mispricing being exploited by the arbitrageur worsens in the short run. Even if General Motors is a *perfect* substitute security for Ford, the arbitrageur still faces the risk that the pessimistic investors who caused Ford to be undervalued in the first place become even more pessimistic, lowering its price even further. Once one has granted the possibility that a price can be different from its fundamental value, then one must also grant the possibility that future price movements will increase the divergence.

Of course, if prices tend to converge toward fundamental value eventually, then arbitrageurs with long horizons care little about noise trader risk: even if the mispricing worsens in the short run, they can just wait out the short term losses, in anticipation of an eventual correction. The reason noise trader risk is important is that many real world arbitrageurs have short, rather than long, horizons. This is because many of the people doing arbitrage – professional portfolio managers – are not managing their own money, but rather managing money for other people. In the words of Shleifer and Vishny (1997), there is a "separation of brains and capital."

This agency feature has important consequences. Investors, lacking the specialized knowledge to evaluate the arbitrageur's strategy, may simply evaluate him based on his returns. If a mispricing that the arbitrageur is trying to exploit worsens in the short run, leading to losses, investors may decide that he is incompetent, and withdraw their funds. Far from being able to wait out the short term losses, the arbitrageur may be forced to liquidate prematurely, just at the time when investment opportunities are at their most attractive. Fear of such premature liquidation makes him act as if his horizon is short.

These problems will only be exacerbated by creditors. After short term losses, creditors, seeing the value of their collateral erode, will call their loans, again triggering premature liquidation.[2]

Implementation Costs

The strategies needed to exploit mispricing are often far from trivial to put in place. Many of the difficulties relate to selling securities short, which is what the arbitrageur must do if he is to avoid fundamental risk. For a large fraction of money managers – pension fund and mutual fund managers in particular – shorting is simply not allowed. A money manager who is allowed to short – a hedge fund manager for example – may still be unable to if the supply of shorts does not meet the demand. Even if he *is* able to short, the arbitrageur cannot be sure that he can continue to borrow the security long enough for the mispricing to correct itself and for him to make money. Should the original owner of the security want it back, the arbitrageur will have to cover his short position by buying the security in the open market at possibly unfavorable terms, a situation known as being "bought in."

Some arbitrage strategies require the purchase or sale of securities in foreign markets. There are often legal restrictions preventing U.S. investors from doing so. Circumventing these restrictions via legal loopholes is costly. Finally, the "implementation costs" category also includes the generic transaction costs arbitrageurs face when implementing strategies, such as commissions or bid-ask spreads.

Model Risk

One final reason why arbitrage may be limited is that even once a mispricing has occurred, arbitrageurs will often still be unsure as to whether it really exists or not. One way to think about this is to imagine that in their search for attractive opportunities, arbitrageurs rely on a model of fundamental value, which tells them, for instance, that the fundamental value of Ford is close to \$20. If noise traders push Ford's price down to \$15, the model will signal a possible mispricing. However, the arbitrageur cannot be *sure* that Ford is mispriced: it is also possible that it is his model that is wrong, and that the stock is in fact correctly priced at \$15. This source of uncertainty, which we label model risk, will also limit the arbitrageur's position.

In contrast, then, to straightforward-sounding textbook arbitrage, real world arbitrage involves a number of risks, which under some conditions will limit arbitrage and allow deviations from fundamental value to persist. To see what these conditions are, consider two cases.

[2]In some cases, noise trader risk can cause problems even when investors manage their own money. If a mispricing is projected to take a long time to close, the annualized expected return from exploiting it may fall below the riskfree rate, making the opportunity unattractive to arbitrageurs.

Suppose first that the mispriced security does *not* have a close substitute security. By definition then, the arbitrageur will be exposed to fundamental risk. In this case, sufficient conditions for arbitrage to be limited are (i) that arbitrageurs are risk averse and (ii) that the fundamental risk is systematic, in that it cannot be diversified by taking many such positions. Condition (i) ensures that the mispricing will not be wiped out by a single arbitrageur taking a large position in the mispriced security. Condition (ii) ensures that the mispricing will not be wiped out by a large number of investors each adding a *small* position in the mispriced security to their current holdings. The presence of noise trader risk, model risk, or implementation costs will only limit arbitrage further.

Even if a perfect substitute does exist, arbitrage can still be limited. The existence of the substitute security immunizes the arbitrageur both from fundamental risk, and from model risk: if two securities with identical cashflows in future states of the world are selling at different prices, he can be completely confident of a mispricing. We can go further and assume that there are no implementation costs, so that only noise trader risk remains. De Long et. al. (1990a) show that noise trader risk is powerful enough, that even with this single form of risk, arbitrage can sometimes be limited. The sufficient conditions are similar to those above. Here arbitrage will be limited if: (i) that arbitrageurs are risk averse *and have short horizons* and (ii) that the noise trader risk is systematic. As before, condition (i) ensures that the mispricing cannot be wiped out by a single, large arbitrageur, while condition (ii) prevents a large number of small investors from exploiting the mispricing.

Attempts to capture other real world issues only make the case for complete arbitrage even more unlikely. For example, there may be other reasons why a large number of different individuals are not able to intervene in an attempt to correct the mispricing. One possibility is that doing the arbitrage requires resources and connections that are only available to a few trained professionals. Alternatively, it may be that there are costs to learning about arbitrage opportunities (Merton, 1987) so that only a handful of people are actually aware of the opportunity at any moment.

So far, we have argued that it is not easy for arbitrageurs like hedge funds to exploit market inefficiencies. However, hedge funds are not the only market participants trying to take advantage of noise traders: firm managers also play this game. If a manager believes that investors are overvaluing his firm's shares, he can benefit the firm by issuing extra shares at attractive prices. The extra supply this generates could potentially push prices back to fundamental value.

Unfortunately, this game is risky for managers, just as it is for hedge funds. In this case, model risk may be particularly important. The manager can rarely be *sure* that investors are overvaluing his firm's shares. If he issues shares, thinking that they are overvalued when in fact they are not, he incurs the costs of deviating from his target capital structure, without getting any benefits in return.

2.3 Evidence

From the theoretical point of view, there is reason to believe that arbitrage is a risky process and therefore that it is only of limited effectiveness. But is there any evidence that arbitrage is limited? In principle, any example of persistent mispricing is immediate evidence of limited arbitrage: if arbitrage were not limited, the mispricing would quickly disappear. The problem is that while many pricing phenomena can be interpreted as deviations from fundamental value, it is only in a few cases that the presence of a mispricing can be established beyond any reasonable doubt. The reason for this is what Fama (1970) dubbed the "joint hypothesis problem." In order to claim that the price of a security differs from its properly discounted future cashflows, one needs a model of "proper" discounting. Any test of mispricing is therefore inevitably a *joint* test of mispricing and of a model of discount rates, making it difficult to provide definitive evidence of inefficiency.

In spite of this difficulty, it turns out that there are a number of financial market phenomena that are almost certainly mispricings, and persistent ones at that. These examples show that arbitrage is indeed limited, and also serve as interesting illustrations of the risks described earlier.

Twin Shares

In 1907, Royal Dutch and Shell Transport, at the time completely independent companies, agreed to merge their interests on a 60:40 basis, while remaining separate entities. Shares of Royal Dutch, which are primarily traded in the U.S. and the Netherlands, are a claim to 60% of the total cashflow of the two companies, while Shell, which trades primarily in the U.K., is a claim to the remaining 40%. If prices equal fundamental value, the value of Royal Dutch equity should always be 1.5 times the value of Shell equity. Remarkably, it isn't.

Figure 1, taken from Froot and Dabora's (1999) analysis of this case, shows the ratio of Royal Dutch equity value to Shell equity value relative to the efficient markets benchmark of 1.5. The picture provides strong evidence of a persistent inefficiency. Moreover, the deviations are not small. Royal Dutch is sometimes 35% underpriced relative to parity, and sometimes 15% overpriced.

This evidence of mispricing is simultaneously evidence of limited arbitrage, and it is not hard to see why arbitrage might be limited in this case. If an arbitrageur wanted to exploit this phenomenon – and several hedge funds, LTCM included, did try to – he would buy the relatively undervalued share and short the other. Table 1 summarizes the risks facing the arbitrageur. Since one share is a good substitute for the other, fundamental risk is nicely hedged: news about fundamentals should affect the two shares equally, leaving the arbitrageur immune. Implementation risk is small: shorting shares of either company is an easy matter. Nor is there any model risk: one of the two shares is almost certainly mispriced. In particular, Froot and Dabora (1999) rule out other possible explanations for the difference in price.

The one risk that remains is noise trader risk. Whatever investor sentiment is causing one share to be undervalued relative to the other could also cause that share to become *even more* undervalued in the short term. The graph shows that this danger is very real: an arbitrageur buying a 10% undervalued Royal Dutch share in March 1983 would have seen it drop still further in value over the next six months. As discussed earlier, when noise trader risk is the only risk facing arbitrageurs, arbitrage will be limited if (i) arbitrageurs are risk averse and have short horizons and (ii) the noise trader risk is systematic, or the arbitrage requires specialized skills, or there are costs to learning about such opportunities. It is very plausible that both (i) and (ii) are true, thus explaining why the mispricing persisted for so long. As we write this in mid-2001, the shares are finally selling at par.

This example is a nice illustration of the distinction between "prices are right" and "no free lunch" discussed in Section 2.1. While prices in this case are *not* right, there are no easy profits for the taking.

ADR's

ADR's are shares of foreign securities held in trust by U.S. financial institutions. Claims on these shares trade in the U.S. In many cases, the ADR of a foreign company trades in New York at a price quite different from the price at which the underlying share trades in its home country.

Once again, this is a clear mispricing: two securities which are claims to the same set of cashflows are trading at different prices. At the same time, it is evidence of limited arbitrage. Why might arbitrage be limited in this case? If the ADR of a Korean company, say, is trading at a premium, an arbitrageur would want to buy the underlying share in Korea and short the ADR. Such a strategy carries substantial implementation costs because there are legal restrictions on foreign ownership of Korean shares, restrictions which are costly to circumvent. However, pricing anomalies exist even in cases without such implementation problems. The reason is noise trader risk: whatever investor sentiment is overpricing the ADR could cause it to become even more overpriced in the short term.

Index Inclusions

Every so often, one of the companies in the S&P 500 leaves the index because of a merger or bankruptcy, and is replaced by another firm. Two early studies of such index inclusions, Harris and Gurel (1986) and Shleifer (1986), document a remarkable fact: when a stock is added to the index, it jumps in price by an average of 3.5%, and much of this jump is permanent. In one dramatic illustration of this phenomenon, when Yahoo! was added to the index, its shares jumped by 24 percent in a single day.

The fact that a stock jumps in value upon inclusion is once again clear evidence of mispricing: the price of the share changes even though its fundamental value does not. The index is a collection of representative companies, and inclusion is not intended to convey any information about the level of a firm's future cashflows nor about their riskiness.

This example of a deviation from fundamental value is also evidence of limited arbitrage. When one thinks about the risks involved in trying to exploit the anomaly, its persistence becomes less surprising. An arbitrageur needs to short the included security and to buy as good a substitute security as he can. This entails considerable fundamental risk because individual stocks rarely have good substitutes. It also carries substantial noise trader risk: whatever caused the initial jump in price – in all likelihood, buying by S&P 500 index funds – may continue, and cause the price to rise still further in the short run.

Wurgler and Zhuravkaya (2000) provide some interesting additional evidence of limited arbitrage. They hypothesize that the jump upon inclusion should be particularly large for those stocks with the worst substitute securities, in other words, for those stocks for which the arbitrage is riskiest. By constructing the best possible substitute portfolio for each included stock, they are able to test this, and find strong support. Their analysis also shows just how hard it is to find good substitute securities for individual stocks. For most regressions of included stock returns on the returns of the best substitute securities, the R^2 is below 25%.

Internet Carve-Outs

In March 2000, 3Com sold 5% of its wholly owned subsidiary Palm Inc. in an initial public offering, retaining ownership of the remaining 95%. After the IPO, a shareholder of 3Com indirectly owned 1.5 shares of Palm. 3Com also announced its intention to spin off the remainder of Palm within 9 months, at which time they would give each 3Com shareholder 1.5 shares of Palm.

At the close of trading on the first day after the IPO, Palm shares stood at $95, putting a lower bound on the value of 3Com of $142. In fact, 3Com's actual price was $81, implying a market valuation of 3Com's substantial businesses outside of Palm of -$60 per share!

This situation surely represents a severe mispricing, and it persisted for several weeks. To exploit it, an arbitrageur could buy one share of 3Com, short 1.5 shares of Palm, and wait for the spin-off, thus earning certain profits at no cost. This strategy entails no fundamental risk, no noise trader risk, and no model risk. Why, then, is arbitrage limited? Lamont and Thaler (2000) who analyze this case in detail, argue that implementation costs play a major role. Many investors who tried to borrow Palm shares to short were either told by their broker that no shares were available, or else were quoted a very high borrowing price. This barrier to shorting was not a legal one, but one that arose endogeneously in the marketplace: such was the demand for shorting Palm, that the supply of Palm shorts was unable to meet it. Arbitrage was therefore limited, and the mispricing persisted.

Palm/3-Com is just one of many "negative stub" situations in which the market value of a company is less that the sum of its publicly traded parts. Mitchell, Pulvino, and Stafford (2000) uncover no less than 70 such examples in the period from 1985 to 2000. In most cases, the parent company did not announce a spin-off, subjecting arbitrageurs to noise trader risk and making it even harder to correct the mispricing.

The reaction of some financial economists to these four examples is to say that they are simply isolated instances with little broad relevance. This may be an overly complacent reaction. The "twin shares" example illustrates that in situations where arbitrageurs face only one form of risk – noise trader risk – securities can become mispriced by 35%. This suggests that if a typical stock trading on the NYSE or NASDAQ becomes subject to investor sentiment, the mispricing could be an order of magnitude larger. Not only would arbitrageurs face noise trader risk in trying to correct the mispricing, but fundamental risk and model risk as well, not to mention possible implementation costs.

As an illustration of how large arbitrage risks might be, consider the meteoric rise of U.S. large stock indices from 1995-2000. To many observers, the S&P 500 and NASDAQ indexes seemed highly overvalued, and yet few dared to act on their hunch. It is not hard to see why. An arbitrageur who shorts the S&P 500 or NASDAQ faces substantial fundamental risk because there is no effective substitute security for value-weighted indexes. He could try going long a small stock index such as the Russell 2000, but he would then still be vulnerable to fundamental news about large stocks that leaves small stocks untouched. There is also noise trader risk: whatever exuberance pushed the S&P 500 and NASDAQ up in the first place could push them up still further in the short run. And finally, there is model risk: an arbitrageur cannot be completely confident that the index is mispriced: perhaps valuations are justified after all, due to lower risk or prospects of higher future earnings.

3 Psychology

The theory of limited arbitrage shows that if irrational traders cause deviations from fundamental value, rational traders will often be powerless to do anything about it. In order to say more about the structure of these deviations, behavioral models often assume a specific form of irrationality. For guidance on this, behavioral economists turn to the extensive experimental evidence compiled by cognitive psychologists on the biases that creep in when people form *beliefs*, and on people's *preferences*. In this section, we summarize the psychology that may be of particular interest to financial economists.

3.1 Beliefs

Overconfidence. Extensive evidence shows that people are overconfident in their judgments. This appears in two guises. First, people are poorly calibrated when estimating probabilities: events they think are certain to occur actually occur only 80% of the time, and events they deem impossible occur 20% of the time. Second, the confidence intervals people assign to their estimates of quantities – the level of the Dow in a year, say – are far too narrow. Their 98% confidence intervals, for example, include the true quantity only 60% of the time.

Optimism and Wishful Thinking. Most people display unrealistically rosy views of their abilities and prospects. Typically over 90 percent of those surveyed think they are above average in such domains as driving skill, ability to get along with people, and sense of humor. They also display a systematic planning fallacy: they predict that tasks (such as writing survey papers) will be completed much sooner than is actually realized.

Representativeness. Kahneman and Tversky (1974) argue that when people try to determine the probability that a data set A was generated by a model B, or that an object A belongs to a class B, they often use the representativeness heuristic. This means that they evaluate the probability by the degree to which A reflects the essential characteristics of B.

Much of the time, representativeness is a helpful heuristic, but it can generate some severe biases. The first is *base rate neglect.* To illustrate, Kahneman and Tversky present this description of a person named Linda:

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

When asked which of "Linda is a bank teller" (statement A) and "Linda is a bank teller and is active in the feminist movement" (statement B) is more likely, subjects typically assign greater probability to B. This is, of course, impossible. Representativeness provides a simple explanation. The description of Linda *sounds* like the description of a feminist – it is representative of a feminist – leading subjects to pick B. Put differently, while Bayes law says that

$$p(\text{statement B}|\text{description}) = \frac{p(\text{description}|\text{statement B})p(\text{statement B})}{p(\text{description})},$$

people apply the law incorrectly, putting too much weight on $p(\text{description}|\text{statement B})$, which captures representativeness, and too little weight on the base rate, $p(\text{statement B})$.

Representativeness also leads to another bias, *sample size neglect.* When judging the likelihood that a data set was generated by a particular model, people do not take into account the size of the sample: after all, a small sample can be just as representative as a large one. Six tosses of a coin resulting in three heads and three tails is as representative of a fair coin as 500 heads and 500 tails are in a total of 1000 tosses.

Sample size neglect means that in cases where people do not initially know the data-generating process, they will tend to infer it too quickly on the basis of too few data points. For instance, they will come to believe that a financial analyst with four good stock picks is talented because four successes are not representative of a bad analyst. It will also generate a "hot hand" phenomenon, whereby sports fans become convinced that a basketball player who has made three shots in a row is on a hot streak, even though there is no evidence of a hot hand in the data. This belief that even small samples will reflect the properties of the parent population is sometimes known as the "law of small numbers" (Rabin, 2001).

In situations where people *do* know the data-generating process in advance, the law of small numbers generates a gambler's fallacy effect. If a fair coin generates five heads in a row, people will say that "tails are due". Since they believe that even a short sample should be representative of the fair coin, there have to be more tails to balance out the large number of heads.

Conservatism. While representativeness leads to an underweighting of base rates, there are situations where base rates are *over*-emphasized relative to sample evidence. In an experiment run by Edwards (1968), there are two urns, one containing 3 blue balls and 7 red ones, and the other containing 7 blue balls and 3 red ones. A random draw of 8 balls from one of the urns (with replacement) yields 8 reds and 4 blues. What is the probability the draw was made from the first urn? While the correct answer is 0.97, most people estimate a number around 0.7, apparently overweighting the base rate of 0.5.

At first sight, the evidence of conservatism appears at odds with representativeness. However, there may be a natural way in which they fit together. It appears that if a data sample is representative of an underlying model, then people overweight the data. However, if the data is not representative of any salient model, people react too little to the data and rely too much on their priors.

Confirmation Bias. Once people have formed a hypothesis, they sometimes misread additional evidence that goes against them as actually being in their favor. They will therefore keep believing in their hypothesis even when contradicted by new data. In a way, this bias is related to conservatism: in both cases, insufficient attention is paid to new data. For example, if people start out believing in the Efficient Markets Hypothesis, they may continue to believe in it long after compelling evidence to the contrary has emerged.

Anchoring. In forming estimates, people often start with some initial, possibly arbitrary value, and then adjust away from it. Experimental evidence shows that the adjustment is often insufficient. Put differently, people "anchor" too much on the initial value.

In one experiment, subjects were asked to estimate the percentage of African countries belonging to the United Nations. More specifically, before giving a percentage, they were asked whether their guess was higher or lower than a randomly generated number between 0 and 100. Their subsequent estimates were significantly affected by the initial random number. Those who were asked to compare their estimate to 10, subsequently estimated 25%, while those who compared to 60, estimated 45%.

Memory Biases. When judging the probability of an event – the likelihood of getting mugged in Chicago, say – people often search their memories for relevant information. This may lead to biases because not all memories are equally retrievable, or "available", in the language of Kahneman and Tversky (1974). More recent events and more salient events – the mugging of a close friend, say – will weigh more heavily and distort the estimate.

Economists are sometimes wary of this body of experimental evidence because they believe (i) that people, through repetition, will learn their way out of biases; (ii) that experts in a field, such as traders in an investment bank, will make fewer errors; and (iii) that with more powerful incentives, the effects will disappear.

While all these factors can attenuate biases to some extent, there is little evidence that they wipe them out altogether. The effect of learning is often muted by errors of application: when the bias is explained, people often understand it, but then immediately proceed to violate it again in specific applications. Expertise, too, is often a hindrance rather than a help: experts, armed with their sophisticated models, have been found to exhibit *more* overconfidence than laymen, particularly when they receive only limited feedback about their predictions.

3.2 Preferences

Prospect Theory

One essential ingredient of any model trying to understand asset prices or trading behavior is an assumption about investor preferences, or about how investors evaluate risky gambles. The vast majority of models assume that investors evaluate gambles according to the expected utility framework, EU henceforth. The theoretical motivation for this goes back to Von Neumann and Morgenstern (1947), VNM henceforth, who show that if preferences satisfy a number of plausible axioms – completeness, transitivity, continuity, and independence – then they can be represented by the expectation of a utility function.

Unfortunately, experimental work in the decades after VNM has shown that people systematically violate EU theory when choosing among risky gambles. In response to this, there has been an explosion of work on so-called non-EU theories, all of them trying to do a better job of matching the experimental evidence. Some of the better known models include weighted-utility theory (Chew and MacCrimmon 1979, Chew 1983), implicit EU (Chew 1989, Dekel 1986), disappointment aversion (Gul 1991), rank-dependent utility theories (Quiggin 1982, Segal 1987, 1989, Yaari 1987), and prospect theory (Kahneman and Tversky 1979, 1992).

Should financial economists be interested in any of these alternatives to expected utility? After all, EU theory may be a good approximation to how people evaluate a risky gamble like the stock market, even if does not explain attitudes towards the kinds of gambles studied in experimental settings. Unfortunately, the difficulty the EU approach has encountered in trying to explain basic facts about the stock market suggests that this is unlikely, and that a closer look at the experimental evidence is warranted. Indeed, recent work in behavioral finance has argued that some of the insights psychologists have drawn from violations of EU are central to understanding a number of financial phenomena.

Of all the non-EU theories, prospect theory may be the most promising for financial applications, and we discuss it in more detail. The reason we focus on this theory is, quite simply, that it is the most successful at capturing the experimental results. In a way, this is not surprising. Most of the other non-EU models are what might be called quasi-normative, in that they try to capture some of the anomalous experimental evidence by slightly weakening the VNM axioms. The difficulty with such models is that in trying to achieve two goals – normative and descriptive – they end up doing an unsatisfactory job at both. In contrast, prospect theory has no aspirations as a normative theory: it simply tries to capture people's attitudes to risky gambles as parsimoniously as possible. Indeed, Kahneman and Tversky (1986) argue convincingly that normative approaches are doomed to failure, because people routinely make choices that are simply impossible to justify on normative grounds, in that they violate dominance or invariance.

Kahneman and Tversky (1979) lay out the original version of prospect theory, designed for gambles with at most two non-zero outcomes. They propose that when offered a gamble

$$(x, p; y, q),$$

to be read as "get outcome x with probability p, outcome y with probability q", people assign it a value of $\pi(p)v(x) + \pi(q)v(y)$, where v and π are shown in Figure 2. When choosing between different gambles, they pick the one with the highest value.

This formulation has a number of key features. First, utility is defined over gains and losses rather than over final wealth positions, an idea first proposed by Markowitz (1952). This fits naturally with the way gambles are often presented and discussed in everyday life. More generally, it is consistent with the way people perceive attributes such as brightness, loudness, or temperature relative to earlier levels, rather than in absolute terms. Kahneman and Tversky (1979) also offer the following violation of EU as evidence that people focus on gains and losses. Subjects are asked:

In addition to whatever you own, you have been given 1000. Choose between

$$\begin{aligned} A = & \quad (1000, 0.5) \\ B = & \quad (500, 1). \end{aligned}$$

B was the more popular choice. The same subjects were then asked:

In addition to whatever you own, you have been given 2000. Choose between

$$\begin{aligned} C &= (-1000, 0.5) \\ D &= (-500). \end{aligned}$$

This time, C was chosen.

Note that the two problems are identical in terms of their final wealth positions and yet people choose differently. The subjects are apparently focusing only on gains and losses. Indeed, when they are not given any information about prior winnings, they again choose B over A and C over D.

The second important feature is the shape of the value function v. It is concave over gains and convex over losses, as suggested by Kahneman and Tversky's finding that[3]

$$(2000, \frac{1}{4}; 4000, \frac{1}{4}) \succ (6000, \frac{1}{4}),$$

and

$$(-6000, \frac{1}{4}) \succ (-4000, \frac{1}{4}; -2000, \frac{1}{4}).$$

In particular, people are risk-seeking over losses.

The v function also has a kink at the origin, indicating a greater sensitivity to losses than to gains, a feature known as *loss aversion*. Loss aversion is introduced to capture aversion to bets of the form:

$$E = (110, \frac{1}{2}; -100, \frac{1}{2}).$$

It may seem surprising that we need to depart from the expected utility framework in order to understand attitudes to gambles as simple as E, but it is nonetheless true. In a remarkable paper, Rabin (2000) shows that if an expected utility maximizer rejects gamble E over some range of wealth levels, then he will also reject

$$(\infty, \frac{1}{2}; -10,000, \frac{1}{2}),$$

an utterly implausible prediction. The intuition is simple: if a utility function defined over final wealth has sufficient local curvature to reject E over a wide range of wealth levels, it

[3]In this section $G_1 \succ G_2$ should be read as "a statistically significant fraction of Kahneman and Tversky's subjects preferred G_1 to G_2."

must be an extraordinarily concave function, making the investor extremely risk averse over large stakes gambles.

The final piece of prospect theory is the nonlinear probability transformation. Small probabilities are overweighted, so that $\pi(p) > p$, in line with KT's finding that

$$(5000, 0.001) \succ (5, 1)$$

and

$$(-5, 1) \succ (-5000, 0.001).$$

Moreover, people are more sensitive to differences in probabilities at higher probability levels. For example, the following pair of choices,

$$(3000, 1) \succ (4000, 0.8; 0, 0.2)$$

and

$$(4000, 0.2; 0, 0.8) \succ (3000, 0.25),$$

which violate EU theory, imply

$$\frac{\pi(0.25)}{\pi(0.2)} < \frac{\pi(1)}{\pi(0.8)}.$$

The intuition is that the 20% jump in probability from 0.8 to 1 is more striking to people than the 20% jump from 0.2 to 0.25. In particular, people place much more weight on outcomes that are certain relative to outcomes that are merely probable, a feature sometimes known as the "certainty effect".

Along with capturing experimental evidence, prospect theory also simultaneously explains preferences for insurance and for buying lottery tickets. Although the concavity of v in the region of gains generally produces risk aversion, for lotteries which offer a small chance of a large gain, the overweighting of small probabilities in Figure 2 dominates, leading to risk-seeking. Along the same lines, while the convexity of v in the region of losses typically leads to risk-seeking, the same overweighting of small probabilities introduces risk aversion over gambles which have a small chance of a large loss.

Based on additional evidence, Tversky and Kahneman (1992) propose a generalization of prospect theory which can be applied to gambles with more than two outcomes. Specifically, if a gamble promises outcome x_i with probability p_i, Tversky and Kahneman (1992) propose that the people assign the gamble the value

$$\sum \pi_i v(x_i)$$

where

$$v = \begin{array}{ll} x^\alpha & \text{if } x \geq 0 \\ -\lambda(-x)^\alpha & \text{if } x < 0 \end{array}$$

and

$$\begin{aligned}\pi_i &= w(P_i) - w(P_i^*)\\ w(P) &= \frac{P^\gamma}{(P^\gamma + (1-P)^\gamma)^{1/\gamma}}.\end{aligned}$$

Here, P_i (P_i^*) is the probability that the gamble will yield an outcome at least as good as (strictly better than) x_i. Tversky and Kahneman (1992) use experimental evidence to estimate $\alpha = 0.88$, $\lambda = 2.25$, and $\gamma = 0.65$. Note that λ is the coefficient of loss aversion, a measure of the relative sensitivity to gains and losses. Over a wide range of experimental contexts λ has been estimated in the neighborhood of 2.

Earlier in this section, we saw how prospect theory could explain why people made different choices in situations with identical final wealth levels. This illustrates an important feature of the theory, namely that it can accommodate the effects of problem description, or *framing*. Such effects are powerful. There are numerous demonstrations of a 30 to 40 percent shift in preferences depending on the wording of a problem. No normative theory of choice can accommodate such behavior since a first principle of rational choice is that choices should be independent of the problem description or representation.

Framing refers to the way a problem is posed for the decision maker. In many actual choice contexts the decision maker also has flexibility in how to think about the problem. For example, suppose that a gambler goes to the race track and wins \$200 in her first bet, but then loses \$50 on her second bet. Does she code the outcome of the second bet as a loss of \$50 or as a reduction in her recently won gain of \$200? In other words, is the utility of the second loss $v(-50)$ or $v(150) - v(200)$? The process by which people formulate such problems for themselves is called *mental accounting* (Thaler, 1999). Mental accounting matters because in prospect theory, v is nonlinear.

One important feature of mental accounting is *narrow framing*, which is the tendency to treat individual gambles separately from other portions of wealth. In other words, when offered a gamble, people often evaluate it as if it is the only gamble they face in the world, rather than merging it with pre-existing bets to see if the new bet is a worthwhile addition.

Redelmeier and Tversky (1991) provide a simple illustration, based on the gamble

$$F = (2000, \frac{1}{2}; -500, \frac{1}{2}).$$

Subjects in their experiment were asked whether they were willing to take this bet; 57% said they would. They were then asked whether they would prefer to play F five times or six times; 70% preferred the six-fold gamble. Finally they were asked:

Suppose that you have played F five times but you don't yet know your wins and losses. Would you play the gamble a sixth time?

60% rejected the opportunity to play a sixth time, reversing their preference from the earlier question. This suggests that some subjects are framing the sixth gamble narrowly, segregating it from the other gambles. Indeed, the 60% rejection level is very similar to the 57% rejection level for the one-off play of F.

Ambiguity Aversion

Our discussion so far has centered on understanding how people act when the outcomes of gambles have known, objective probabilities. In reality, probabilities are rarely objectively known. To handle these situations, Savage (1964) develops a counterpart to expected utility known as subjective expected utility, SEU henceforth. Under certain axioms, preferences can be represented by the expectation of a utility function, this time weighted by the individual's subjective probability assessment.

Experimental work in the last few decades has been as unkind to SEU as it was to EU. The violations this time are of a different nature, but they may be just as relevant for financial economists.

The classic experiment was described by Ellsberg (1961). Suppose that there are two urns, 1 and 2. Urn 2 contains a total of 100 balls, 50 red and 50 blue. Urn 1 also contains 100 balls, again a mix of red and blue, but the subject does not know the proportion of each.

Subjects are then asked to choose one of the following two gambles, each of which involves a possible payment of \$100, depending on the color of a ball drawn at random from the relevant urn

$$
\begin{aligned}
a_1 &: \quad \text{a ball is drawn from Urn 1, \$100 if red, \$0 if blue} \\
a_2 &: \quad \text{a ball is drawn from Urn 2, \$100 if red, \$0 if blue.}
\end{aligned}
$$

Subjects are then asked to choose between following two gambles:

$$
\begin{aligned}
b_1 &: \quad \text{a ball is drawn from Urn 1, \$100 if blue, \$0 if red} \\
b_2 &: \quad \text{a ball is drawn from Urn 2, \$100 if blue, \$0 if red.}
\end{aligned}
$$

a_2 is typically preferred to a_1, while b_2 is chosen over b_1. These choices are inconsistent with SEU: the choice of a_2 implies a subjective probability that *fewer* than 50% of the balls in Urn 1 are red, while the choice of b_2 implies the opposite.

This experiment suggests that people dislike subjective, or vague uncertainty more than they dislike objective uncertainty, a finding often labelled "ambiguity aversion". Ambiguity can be defined as a situation where information that could be known, is not – the proportion of red and blue balls, in our example.

Subsequent work has uncovered reliable evidence of ambiguity aversion in more realistic settings where people bet on events such as the outcome of a football match. Ambiguity aversion is particularly strong in cases where people feel that their competence in assessing the relevant probabilities is low (Heath, Tversky 1991). This effect can be strengthened further by reminding subjects of their incompetence, either through comparison with other bets in which they have more expertise, or by comparison with other people who are more qualified to evaluate the bet (Fox, Tversky 1995).

4 Application: The Aggregate Stock Market

Researchers studying the aggregate stock market have identified a number of interesting stylized facts about its behavior. Three of the most striking are:

(i) the *equity premium*: Using annual data from 1871-1993, Campbell and Cochrane (1999) report the average log return on the S&P 500 index to be 3.9% higher than the average return on short term commercial paper.

(ii) *volatility*: In the same data set, the standard deviation of excess log returns on the S&P 500 is 18%, while the standard deviation of the log price-dividend ratio is 0.27.

(iii) *predictability*: Based on monthly, equal-weighted, real NYSE returns from 1941-1986, Fama and French (1988) show that the dividend-price ratio is able to explain 27% of the variation of cumulative stock returns over the subsequent four years.

All three of these facts have labelled "puzzles". Fact (i) has been known as the equity premium puzzle since the work of Mehra and Prescott (1985) (see also Hansen and Singleton, 1983). Campbell (2000) calls (ii) the volatility puzzle, and we refer to (iii) as the predictability puzzle. The reason they are referred to as puzzles is that they are hard to rationalize in a simple consumption-based model.

To see this, consider the following endowment economy, which we come back to a number of times in this section. There are an infinite number of identical investors, and two assets: a riskfree asset in zero net supply, with gross return $R_{f,t}$ between time t and $t+1$, and a risky asset – the stock market – in fixed positive supply, with gross return R_{t+1} between time t and $t+1$. The stock market is a claim to a perishable stream of dividends $\{D_t\}$, where

$$\frac{D_{t+1}}{D_t} = e^{g_D + \sigma_D \varepsilon_{t+1}}.$$

Each period's dividend can be thought of as one component of a consumption endowment C_t, where

$$\frac{C_{t+1}}{C_t} = e^{g_C + \sigma_C \eta_{t+1}},$$

with

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \omega \\ \omega & 1 \end{pmatrix}\right), \text{ i.i.d. over time.}$$

Investors choose consumption C_t and an allocation S_t to the risky asset to maximize

$$E_0 \sum_{t=0}^{\infty} \rho^t \frac{C_t^{1-\gamma}}{1-\gamma}, \tag{1}$$

subject to the standard budget constraint. Based on the Euler equation of optimality,

$$1 = \rho E_t \left[\left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} R_{t+1} \right],$$

it is straightforward to compute the properties of stock returns in this economy.[4] For the parameter values in Table 2, the average return on the stock market is only 0.1% higher, not 3.9% higher, than the average return on T-Bills. The standard deviation of stock returns is only 12%, not 18%, and the price-dividend ratio is completely constant. Of course, this means that the dividend-price ratio has no forecast power for future returns.

It is helpful to recall the intuition for these results. In an economy with power utility investors, the equity premium is determined by risk aversion γ and by risk, which for these investors is measured by the covariance of stock returns and consumption growth. Since consumption growth is very smooth in the data, this covariance is very low, thus predicting a very low equity premium. Stocks simply do not appear risky to investors with the preferences in (1), and therefore do not warrant a large premium.[5]

To understand the volatility puzzle, note that in the simple economy described above, both discount rates and expected dividend growth are constant over time. A direct application of the present value formula immediately implies that the price-dividend ratio is constant. Since

$$R_{t+1} = \frac{1 + P_{t+1}/D_{t+1}}{P_t/D_t} \frac{D_{t+1}}{D_t},$$

it follows that

$$r_{t+1} - E_t(r_{t+1}) = d_{t+1} - E_t(d_{t+1}),$$

where lower case letters indicate log variables. The standard deviation of returns will therefore only be as high as the standard deviation of dividend growth, namely 12%.

The particular volatility puzzle seen here illustrates a more general point, first made by Shiller (1981) and Le Roy and Porter (1981), namely that it is difficult to explain the historical volatility of stock prices with *any* model in which investors make rational forecasts

[4]Full details are in the Appendix (to be added).

[5]Of course, the equity premium predicted by the model can be increased by using higher values of γ. However, the model would then also predict a counterfactually high riskfree rate, a problem known as the riskfree rate puzzle.

of future cashflows and in which discount rates are constant. The intuition is simple: if discount rates are constant, a high price-dividend ratio can only be due to expectations of high dividend growth. If these expectations are to be considered rational, however, it must be that high price-dividend ratios actually *are* on average followed by unusually high cashflow growth. Unfortunately, Campbell (2000) reports that price-dividend ratios are *not* reliable forecasters of dividend growth, neither in the U.S. nor in most international markets.

Shiller and Le Roy and Porter's results shocked the profession when they first appeared. At the time, most economists felt that discount rates *were* constant over time, apparently implying that stock market volatility could only be fully explained by appealing to investor irrationality. Today, it is well known that rational variation in discount rates can help explain the volatility puzzle, although we argue later that models with irrational beliefs may be a more plausible way of thinking about the data.

Behavioral finance has made a number of advances in understanding the three puzzles singled out at the start of this section. We first discuss the equity premium puzzle, and then turn to the volatility puzzle. We do not consider the predictability puzzle separately, because in any model with a stationary price-dividend ratio, a resolution of the volatility puzzle is simultaneously a resolution of the predictability puzzle. To see this, note that if the price-dividend ratio $\frac{P}{D}$ is unusually high, the only way it can return to its average level is if cashflows go up, or if prices fall. As we noted earlier, high price-dividend ratios are *not* on average followed by high cashflows, which means that they must predict lower returns, exactly the predictability puzzle.

4.1 The Equity Premium Puzzle

To date, behavioral finance has pursued two approaches to the equity premium puzzle. Both are based on preferences: one relies on prospect theory, the other on ambiguity aversion. Both approaches argue that for reasons not captured by the preferences in (1), investors find stocks unappealing and are unwilling to allocate much of their wealth to them. Put differently, they are only willing to hold the market supply of equity in return for a very substantial equity premium.

Prospect Theory

The first paper to apply prospect theory to finance was Benartzi and Thaler (1995), BT henceforth. They investigate how an investor with prospect theory-type preferences allocates wealth between T-Bills and the stock market. In particular, they suppose that he maximizes

$$E_{\pi}\ v[(1-\omega)R_{f,t}+\omega R_{t+1}-1], \tag{2}$$

where π and v are defined in Section 3.2, and $R_{f,t}$ and R_{t+1} are the gross returns on T-Bills and the stock market between t and $t+1$ respectively, distributed according to their historical distribution. The investor's control variable is ω, the fraction of his financial wealth allocated to stocks.

In asserting that (2) is the relevant portfolio problem for an investor with prospect-type preferences, BT are assuming that the "gains" and "losses" of prospect theory refer to *changes in financial wealth*. This can be thought of as a narrow framing assumption: even if investors have many forms of wealth, both financial and nonfinancial, they still get utility from changes in the value of the specific component of their wealth made up by their financial holdings.

BT also need to make an assumption about the length of the time interval $[t, t+1]$ over which gains and losses are measured. To see why, compare two investors, Nick who calculates the gains and losses in his portfolio every day, and Dick who only looks at his portfolio once per decade. Since, on a daily basis, stocks go down in value almost as often as they go up, Nick's loss aversion will make stocks appear very unattractive to him. In contrast, loss aversion will not have much effect on Dick's perception of stocks since at ten year horizons stocks offer only a small risk of losing money.

In the absence of direct evidence on how often people evaluate their portfolios, BT ask the question: how often would people have to be evaluating their portfolios in order to make stocks and bonds equally attractive? The answer they obtain is roughly once per year, the exact answer depending on whether stocks are compared to bonds or T-bills, and on whether the analysis is done in real or nominal dollars. BT argue that this result seems plausible, since we receive our most comprehensive mutual fund reports once a year, and do our taxes once a year, suggesting that gains and losses are probably most naturally expressed as *annual* changes in value. BT call the combination of loss aversion and frequent evaluations *myopic loss aversion*.

BT's results suggest that loss aversion over annual changes in financial wealth may be one way of understanding why investors are happy to hold the market supply of stocks even when they know that stocks offer a sizeable equity premium. It is important to note that these findings are indeed only suggestive. The equity premium puzzle is in large part a consumption puzzle: given the low volatility of consumption growth, why are investors so reluctant to buy an asset, stocks, especially when that asset's covariance with consumption growth is so low? Since BT do not consider an intertemporal model with consumption choice, they cannot address this issue directly.

To see if prospect theory can in fact help with the equity premium puzzle, Barberis, Huang and Santos (2001), BHS henceforth, make a first attempt at building it into an equilibrium model of stock returns. A simple version of their model, an extension of which we consider in a later section, examines an economy with the same structure as the one described at the start of this section, but in which investors have the preferences

$$E_0 \sum_{t=0}^{\infty} \left[\rho^t \frac{C_t^{1-\gamma}}{1-\gamma} + b_0 \overline{C}_t^{-\gamma} v(X_{t+1}) \right]. \tag{3}$$

The investor gets utility from consumption, but over and above that, he gets utility from changes in the value of his holdings of the risky asset between t and $t+1$, denoted here by X_{t+1}. BHS therefore follow BT in making the narrow framing assumption that the investor derives utility from changes in the value of a specific component of his wealth. They also follow BT in defining the unit of time to be a year, so that gains and losses are measured annually.

The utility from these gains and losses is determined by $\widehat{v}$ where[6]

$$\widehat{v}(X) = \begin{cases} X & \text{for} \quad X \geq 0 \\ 2.25X & \phantom{\text{for}} \quad X < 0 \end{cases}. \tag{4}$$

The 2.25 factor comes from Tversky and Kahneman's (1992) experimental study of attitudes towards gambles. This specification is simpler than the one used by BT, v. It captures loss aversion, but ignores other elements of prospect theory, such as the concavity (convexity) over gains (losses) and the probability transformation. In part this is because it is difficult to incorporate all these features into a fully dynamic framework; but also, it is based on BT's observation that it is mainly loss aversion that drives their partial equililbrium results.

BHS show that loss aversion can indeed provide a partial rationalization of the high Sharpe ratio on the aggregate stock market. However, how much of the Sharpe ratio it can explain depends heavily on the importance of the second source of utility in (3), or in short, on b_0. As a way of thinking about this parameter, BHS note that when $b_0 = 0.7$, the psychological pain of losing \$100 in the stock market, captured by the second term, is roughly equal to the consumption-based pain of having to consume \$100 less, captured by the first term. For this b_0, the Sharpe ratio of the risky asset is 0.11, about a third of its historical value.

BT and BHS both assume that investors frame gains and losses narrowly not only in a cross-sectional sense, but also in a temporal sense: even though they have long horizons, they still pay attention to annual gains and losses, perhaps because they are presented with annual feedback on their investment performance. Thaler, Tversky, Kahneman, and Knetsch (1997) provide an experimental test of the idea that the way information is presented affects the frame people adopt in their decision-making. Subjects are asked to imagine that they

[6] The $b_0 \overline{C}_t^{-\gamma}$ coefficient on the loss aversion term is a scaling factor which ensures that risk premia in the economy remain stationary even as aggregate wealth increases over time. It involves per capita consumption $\overline{C}_t$ which is exogeneous to the investor, and so does not affect the intuition of the model. The constant b_0 controls the importance of the loss aversion term in the investor's preferences; setting $b_0 = 0$ reduces the model to the much studied case of power utility over consumption.

are portfolio managers for a small college endowment. One group of subjects – Group I, say – is shown monthly observations on two funds, Fund A and Fund B. Returns on Fund A (B) are drawn from a normal distribution calibrated to mimic bond (stock) returns as closely as possible, although subjects are not given this information. After each monthly observation, subjects are asked to allocate their portfolio between the two funds over the next month. They are then shown the realized returns over that month, and asked to allocate once again.

A second group of investors – Group II – is shown exactly the same series of returns, except that it is aggregated at the annual level; in other words, these subjects do not see the monthly fund fluctuations, but only cumulative annual returns. After each annual observation, they are asked to allocate their portfolio between the two funds over the next year.

A final group of investors – Group III – is shown exactly the same data, this time aggregated at the five-year level, and they too are asked to allocate their portfolio after each observation.

After going through a total of 200 months worth of observations, each group is asked to make one final portfolio allocation, which is to apply over the next 400 months. TTKS find that the average final allocation chosen by subjects in Group I is much lower than that chosen by people in Groups II and III. This result is consistent with the idea that people code gains and losses based on how information is presented to them. Subjects in Group I see monthly observations and hence more frequent losses. If they adopt the monthly distribution as a frame, they will be more wary of stocks and will allocate less to them.

Ambiguity Aversion

In Section 3, we presented the Ellsberg paradox as evidence that people dislike ambiguity, or situations where they are not sure what the probability distribution of a gamble is. This is potentially very relevant for finance, as investors are often uncertain about the distribution of a stock's return.

Following the work of Ellsberg, many models of how people react to ambiguity have been proposed; Camerer and Weber (1992) provide a comprehensive review. One of the more popular approaches is to suppose that when faced with ambiguity, people entertain a range of possible probability distributions and act to maximize the minimum expected utility under any candidate distribution. In effect, people behave as if playing a game against a malevolent opponent who picks the actual distribution of the gamble so as to leave them as worse off as possible. Such a decision rule was first axiomatized by Gilboa and Schmeidler (1989). Epstein and Wang (1994) showed how such an approach could be incorporated into a dynamic asset pricing model, although they did not try to assess the quantitative implications of ambiguity aversion for asset prices.

Quantitative implications *have* been derived using a closely related framework known as robust control. In this approach, the agent has a reference probability distribution in mind, but wants to ensure that his decisions are good ones even if the reference model is mispecified to some extent. Here too, the agent essentially tries to guard against a "worst-case" misspecification. Anderson, Hansen, and Sargent (1998) show how such a framework can be used for portfolio choice and pricing problems, even when state equations and objective functions are nonlinear.

Maenhout (1999) applies the AHS framework to the specific issue of the equity premium. He shows that if investors are concerned that their model of stock returns is misspecified, they will charge a substantially higher equity premium as compensation for the perceived ambiguity in the probability distribution. He notes, however, that to explain the full 3.9% equity premium requires an unreasonably high concern about misspecification. At best then, ambiguity aversion is only a partial resolution of the equity premium puzzle.

4.2 The Volatility Puzzle

In thinking about the volatility puzzle, it is useful to consider a version of the present value formula originally derived by Campbell and Shiller (1988). Starting from

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t},$$

where P_t is the value of the stock market at time t, they use a log-linear approximation to show that the log price-dividend ratio can be written

$$p_t - d_t = E_t \sum_{j=0}^{\infty} \rho^t \Delta d_{t+1+j} - E_t \sum_{j=0}^{\infty} \rho^t r_{t+1+j} + \lim_{j \to \infty} \rho^j (p_{t+j} - d_t) + \text{const}, \tag{5}$$

where lower case letters represent log variables – $p_t = \log P_t$, for example – and where $\Delta d_{t+1} = d_{t+1} - d_t$.

This equation is best thought of as an accounting identity: if prices are high relative to current dividends, this must be due either to investors' expecting high future dividend growth or to their expecting low future returns. The essence of the volatility puzzle uncovered by Shiller (1981) and Le Roy and Porter (1981) is that the empirical volatility of the price-dividend ratio – the left-hand side – cannot easily be explained by models which rely only on rational variation in expected dividend growth rates, the first term on the right hand side.

Before turning to behavioral work on the volatility puzzle, it is worth using (5) as a guide to the rational stories that have been proposed. Since the first term on the right hand side cannot be the basis of a successful rational story, researchers have turned to the second term: rational variation in discount rates. Such a story cannot be about rational variation in *riskless* rate forecasts, because then high price-dividend ratios would need to forecast low interest rates on average, which they do not in the data. Instead, one has to tell a story about rational variation in the risk premium. Once again, researchers are limited as to the avenues they can pursue: it is difficult to tell a story based on risk perception, because movements in expected returns do not match up well with changes in standard measures of risk (French, Schwert, Stambaugh, 1987). The only story one can tell is one of changing risk aversion, and this is the idea behind the Campbell and Cochrane (1999) model of aggregate stock market behavior. They propose a habit formation framework in which changes in consumption relative to habit lead to changes in risk aversion and hence variation in price-dividend ratios over and above that due to changing cashflow forecasts.

Some rational approaches have focused on the third term in (5) – in other words, they are stories about rational bubbles. However, a number of papers, most recently Santos and Woodford (1997), show that the conditions under which rational bubbles can survive are extremely restrictive.[7]

We group the behavioral approaches to the volatility puzzle by whether they focus on beliefs or preferences. In general, belief-based stories center on the first or third terms on the right-hand side of (5) – changing forecasts of future cashflows or of future returns – while preference-based stories typically revolve around the second term, namely changing discount rates.

Beliefs

One possible story is that investors believe that the mean dividend growth rate is more variable than it actually is. When they see a surge in dividends, they are too quick to believe that the mean dividend growth rate has increased. Their exuberance pushes prices up relative to dividends.

A story of this kind can be derived as a direct application of representativeness and in particular, of the version of representativeness known as the law of small numbers, whereby people expect even short samples to reflect the properties of the parent population. If the investor sees many periods of good earnings, the law of small numbers leads him to believe that earnings growth has gone up, and hence earnings will continue to be high in the future. After all, the earnings growth rate cannot be "average". If it were, then according the to law of small numbers, earnings should *appear* average, even in short samples: some good earnings news, some bad earnings news, but not several good pieces of news in a row.

Another belief-based story relies more on private, rather than public information, and in particular, on overconfidence about private information. Suppose that an investor has seen public information about the economy, and has formed a prior about future cashflow growth. He then does some research on his own and become overconfident about the information he gathers: he overestimates its accuracy and puts too much weight on it relative to his prior.

[7]Brunnermeier (2001) provides a comprehensive review of this literature.

If the private information is positive, he will push prices up too high relative to current dividends.

Price-dividend ratios might also be excessively volatile because investors extrapolate past *returns* too far into the future when forming expectations of future returns. Such a story might again be based on representativeness and the law of small numbers. The same argument for why investors might extrapolate past cashflows too far into the future can be applied here to explain why they might extrapolate past returns too far into the future.

Modigliani and Cohn (1979) and more recently, Ritter and Warr (2000), have argued that part of the variation in price-dividend ratios may be due to investors confusing real and nominal quantities when forecasting future cashflows. The value of the stock market can be determined by discounted real cashflows at real rates, or nominal cashflows at nominal rates. At times of especially high or especially low inflation though, it is possible that some investors mistakenly discount *real* cashflows at *nominal* rates. If inflation increases, so will the nominal discount rate. If investors then discount the *same* set of cashflows at this higher rate, they will push the value of the stock market down. Of course, this calculation is incorrect: the same inflation which pushes up the discount rate should also push up future cashflows. On net, inflation should have little effect on market value. Such real vs. nominal confusion seems particularly relevant to understanding the low market valuations during the high inflation years of the 1970's, as well as the high market valuations during the low inflation 1990's.

Preferences

Barberis, Huang and Santos (2001) show that a straightforward extension of the version of their model discussed in Section 4.1 can explain both the equity premium and volatility puzzles. To do this, they appeal to experimental evidence about dynamic aspects of loss aversion. This evidence suggests that the degree of loss aversion is not the same in all circumstances but depends on prior gains and losses. In particular, Thaler and Johnson (1990) find that after prior gains, subjects take on gambles they normally do not, and that after prior losses, they refuse gambles that they normally accept. The first finding is sometimes known as the "house money" effect, reflecting gamblers' increasing willingness to bet when ahead. One interpretation of this evidence is that losses are less painful after prior gains because they are cushioned by those gains. However, after being burned by a painful loss, people may become more wary of additional setbacks.

To capture these ideas, Barberis, Huang, and Santos (2001) modify the utility function in (3) to

$$E_0 \sum_{t=0}^{\infty} \left[\rho^t \frac{C_t^{1-\gamma}}{1-\gamma} + b_0 \overline{C}_t^{-\gamma} \tilde{v}(X_{t+1}, z_t) \right].$$

Here, z_t is a state variable that tracks past gains and losses on the stock market. The function $\tilde{v}$ is a piecewise linear function similar in form to v, defined in (4). However, the investors' sensitivity to losses is no longer constant at 2.25, but is determined by z_t, in a way that reflects the experimental evidence described above.

A model of this kind can help explain the volatility puzzle. Suppose that there is some good cashflow news. This pushes the stock market up, creating a cushion of prior gains for investors, who now become less risk averse. They therefore discount future cashflows at a lower rate, pushing prices up still further relative to current dividends.

5 Application: The Cross-section of Average Returns

While the behavior of the aggregate stock market is not easy to understand from the rational point of view, promising rational models have nonetheless been developed and can be tested against behavioral alternatives. Empirical studies of the behavior of *individual* stocks have unearthed a set of facts which is altogether more frustrating for the rational paradigm. Many of these facts are about the *cross-section* of average returns: they document that one group of stocks earns higher average returns than another. These facts have come to be known as "anomalies" because they cannot be explained by the simplest and most intuitive model of risk and return in the financial economist's toolkit, the Capital Asset Pricing Model, or CAPM. There is now a growing sentiment, both in the academic profession and among practitioners that some of the cross-sectional evidence is anomalous not only relative to the CAPM but possibly relative to *any* rational model of risk and return.

We now outline some of the more salient findings in this literature and then consider some of the rational and behavioral approaches in more detail.

The Size Premium

This anomaly was first documented by Banz (1981). We report the more recent findings of Fama and French (1992). Every year from 1963 to 1990, Fama and French group all stocks traded on the NYSE, AMEX, and NASDAQ into deciles based on their market capitalization, and then measure the average return of each decile over the next year. They find that for this sample period, the annual return of the smallest stock decile is 0.74% per month higher than the average return of the largest stock decile. This is certainly an anomaly relative to the CAPM: while stocks in the smallest decile do have higher β's, the difference in risk is not nearly enough to explain the difference in average returns.

Long-term Reversals

Every three years from 1926 to 1982, De Bondt and Thaler (1985) rank all stocks traded on the NYSE by their prior three year cumulative return and form two portfolios: a "winner" portfolio of the 35 stocks with the best prior record and a "loser" portfolio of the 35 worst performers. They then measure the average return of these two portfolios over the three years subsequent to their formation. They find that over the whole sample period, the average annual return of the loser portfolio is higher than the average return of the winner portfolio by about 8% per year.

The Predictive Power of Scaled-price Ratios

These anomalies, which are about the cross-sectional predictive power of variables like the book-to-market (B/M) and earnings-to-price (E/P) ratios, where some measure of fundamentals is scaled by price, were first noted by Basu (1983) and Rosenberg, Reid, and Lanstein (1985). We report Fama and French's (1992) more recent evidence.

Every year, from 1963 to 1990, Fama and French group all stocks traded on the NYSE, AMEX, and NASDAQ into deciles based on their book-to-market ratio, and measure the average return of each decile over the next year. They find that the average return of the highest-B/M-ratio decile, containing so called "value" stocks, is 1.53% per month higher than the average return on the lowest-B/M-ratio decile, "growth" or "glamor" stocks, a difference much higher than can be explained through differences in β between the two portfolios. Repeating the calculations with the earnings-price ratio as the ranking measure produces a difference of 0.68% per month between the two extreme decile portfolios, again an anomalous result.

Momentum

Every month from January 1963 to December 1989, Jegadeesh and Titman (1993) group all stocks traded on the NYSE into deciles based on their prior six month return and compute average returns of each decile over the six months after portfolio formation. They find that the decile of biggest prior winners outperforms the decile of biggest prior losers by an average of 10% on an annual basis.

Comparing this result to De Bondt and Thaler's (1985) study of prior winners and losers illustrates the crucial role played by the length of the prior ranking period. In one case, prior winners continue to win; in the other, they perform poorly. A challenge to both behavioral and rational approaches is to explain why extending the formation period switches the result in this way.

There is some evidence that tax-loss selling creates seasonal variation in the momentum effect. Stocks with poor performance during the year may later be subject to selling by investors keen to realize losses that can offset capital gains elsewhere. This selling pressure means that prior losers continue to lose, enhancing the momentum effect. At the turn of the year, though, the selling pressure eases off, allowing prior losers to rebound and weakening the momentum effect. A careful analysis by Grinblatt and Moskowitz (1999) finds that on net, tax-loss selling may explain part of the momentum effect, but by no means all of it.

In any case, while selling a stock for tax purposes is rational, a model of predictable price movements based on such behavior is not. Roll (1983) calls such explanations "stupid" since investors would have to be stupid not to buy in December if prices were going to increase in January.

A number of studies have examined stock returns following important corporate announcements, a type of analysis known as an event study.

Event Studies of Earnings Announcements

Every quarter from 1974 to 1986, Bernard and Thomas (1989) group all stocks traded on the NYSE and AMEX into deciles based on the size of the surprise in their most recent earnings announcement. "Surprise" is measured relative to a simple random walk model of earnings. They find that on average, over the 60 days after the earnings announcement, the decile of stocks with surprisingly good news outperforms the decile with surprisingly bad news by an average of about 4%, a phenomenon known as post-earnings announcement drift. Once again, this difference in returns is not explained by differences in β between the two portfolios. A later study by Chan, Jegadeesh, and Lakonishok (1996) measures surprise in other ways – relative to analyst expectations, and by the stock price reaction to the news – and obtains similar results.

Event Studies of Dividend Initiations and Omissions

Michaely, Thaler, and Womack (1998) study firms which have announced initiation or omission of a dividend payment. They find that on average, the shares of firms initiating (omitting) dividends outperform (underperform) a control group by a substantial margin over the year after the announcement.

Event Studies of Stock Repurchases

Ikenberry, Lakonishok, and Vermaelen (1995) study firms which have announced a share repurchase between 1980 and 1990. They find that on average, the shares of these firms outperform a control group by a substantial margin over the four year period following the announcement.

Event Studies of Primary and Secondary Offerings

Loughran and Ritter (1995) study firms which have undertaken primary or secondary equity offerings. They find that the average return of shares of these firms over the five year period after the issuance is markedly below the average return of a control group.

Long-term event studies like the last three analyses summarized above raise some thorny statistical problems. Barber and Lyon (1997), Barber, Lyon, and Tsai (1999), Brav (2000), Fama (1998), Loughran and Ritter (2000), and Mitchell and Stafford (2001) are just a few of the papers that discuss this topic. Cross-sectional correlation is an important issue: if one firm announces a share repurchase shortly after another firm does, their four-year post event returns will overlap and cannot be considered independent. Although the problem is an obvious one, it is not easy to deal with effectively; some recent attempts to do so suggest that the anomalous evidence in the event studies on dividend announcements, repurchase announcements, and equity offerings is statistically weaker than initially appeared, although how much weaker remains controversial.

A more general concern with *all* the above empirical evidence is data-mining. After all, if one sorts and ranks stocks in enough different ways, one is bound to discover striking – but completely spurious – cross-sectional differences in average returns.

A first response to the data-mining critique is to note that the above studies do not use the kind of obscure firm characteristics or marginal corporate announcements that would suggest data-mining. Indeed, it is hard to think of an important class of corporate announcements that has *not* been associated with a claim about anomalous post-event returns. A more direct check is to perform out-of-sample tests. Interestingly, a good deal of the above evidence *has* been replicated in other data sets. Fama, French, and Davis (2000) show that there is a value premium in the subsample of U.S. data that precedes the data set used by Fama and French in their 1992 study, while Fama and French (1998) document the presence of a value premium in international stock markets. Rouwenhourst (1997) shows that the momentum effect is alive and well in international stock market data.

If the empirical results are taken at face value, then the challenge to the rational paradigm is to show that the above cross-sectional evidence emerges naturally from a model of the economy in which rational investors maximize a normatively acceptable utility function. In special cases, models of this form reduce to the CAPM, and we know that this does not explain the evidence. More generally, rational models predict a multifactor pricing structure,

$$\overline{r}_i - r_f = \beta_{i,1}(\overline{F}_1 - r_f) + \ldots + \beta_{i,K}(\overline{F}_K - r_f),$$

where the loadings $\beta_{i,k}$ come from a time series regression,

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_{i,1}(F_{1,t} - r_{f,t}) + \ldots + \beta_{i,K}(F_{K,t} - r_{f,t}) + \varepsilon_{i,t}.$$

To date, it has proved difficult to derive a multifactor model which explains the cross-sectional evidence, although this remains a major research direction.

Alternatively, one can skip the step of *deriving* a factor model, and simply try a specific model to see how it does. This is the approach of Fama and French (1993, 1996). They show that a specific three factor model can explain the cross-sectional variation in portfolios formed on size and book-to-market rankings, with an R^2 of over 90%. Their factors are the return on the market portfolio, the return of a portfolio of small stocks minus the return on a portfolio of large stocks – the "size" factor – and the return on a portfolio of value stocks minus the return on a portfolio of growth stocks – the "book-to-market" factor.

The high R^2's obtained by Fama and French (1996) are not necessarily cause for celebration. As Roll (1977) emphasizes, in any specific sample, it is possible to construct a *one* factor model that produces an R^2 of 100%. To be fair, the Fama and French (1993, 1996) factors are not the result of a data-mining exercise. They begin by pointing out that small stocks and value stocks move together. The size and book-to-market factors are an attempt to isolate these common factors in small and value stocks, and their three factor model is motivated by the idea that this comovement is a systematic risk that is priced in equilibrium.

Fama and French (1996) themselves admit that their results can only have their full impact once it is explained what it is about investor preferences and the structure of the economy that leads people to price assets according to their model.

One general feature of the rational approach is that it is loadings or betas, and not firm characteristics that determine average returns. For example, a risk-based approach would argue that value stocks earn high returns not because they have high book-to-market ratios, but because such stocks happen to have a high loading on the book-to-market factor. Daniel and Titman (1996) cast doubt on this specific prediction by performing double sorts of stocks by both book to market ratios and loadings on book-to-market factors. In particular, they show that stocks with different loadings but the same book-to-market ratio do *not* differ in their average returns. These results appear quite damaging to rational approach. However, using a longer data set and with a difference in methodology, Fama, French, and Davis (2000) claim to reverse Daniel and Titman's findings. We expect further developments on this controversial front.

More generally, rational approaches to the cross-sectional evidence face a number of other obstacles. First, rational models typically measure risk as covariance of returns with marginal utility of consumption. Stocks are risky if they fail to pay out at times of high marginal utility – in "bad" times – and instead pay out when marginal utility is low – in "good" times. The problem is that in the above findings, there is little evidence that the portfolios with anomalously *high* average returns do poorly in bad times, whatever plausible measure of bad times is used. For example, Lakonishok, Shleifer, and Vishny (1994) show that value stocks do *well* when the economy is in recession. Similarly, De Bondt and Thaler (1987) found that their loser stocks had higher betas than winners in up markets and lower betas in down markets – an attractive combination.

Second, some of the portfolios in the above studies – the decile of stocks with the lowest book-to-market ratios for example – earn average returns below the riskfree rate. It is hard to see why a rational investor would willingly accept a lower return than the T-Bill rate on a volatile portfolio.

Finally, in some of the examples given above, it is not just that one portfolio outperforms another on average. In some cases, the outperformance is present in almost every period of the sample. For example, in Bernard and Thomas' (1989) study, firms with surprisingly good earnings outperform those with surprisingly poor earnings in 46 out of the 50 quarters studied. It is not easy to see any risk here than might justify the outperformance.

There are a number of behavioral models which try to explain some of the above phenomena. We classify them based on whether their mechanism centers on beliefs or on preferences.

5.1 Belief-based Models

Barberis, Shleifer, and Vishny (1998), BSV henceforth, argue that much of the above evidence is the result of systematic errors that investors make when they use public information to form expectations of future cashflows. They consider a model with a representative risk neutral investor in which the true earnings process for all assets is a random walk. Investors, however, do not use the random walk model to forecast future earnings. They think that at any time, earnings are being generated by one of two regimes: a "mean-reverting" regime, in which earnings are more mean-reverting than in reality, and a "trending" regime in which earnings trend more than in reality. The investor believes that the regime generating earnings changes exogeneously over time and sees his task as trying to figure out which of the two regimes is currently generating earnings.

BSV argue that their model captures two important updating biases discussed in Section 3: conservatism and representativeness, and in particular, the version of representativeness known as the law of small numbers, whereby people expect even short samples to reflect the properties of the parent population. If the investor sees many periods of good earnings, the law of small numbers leads him to believe that this is a firm with particularly high earnings growth, and hence to forecast high earnings in the future. After all, the firm cannot be "average". If it were, then according the to law of small numbers, its earnings should *appear* average, even in short samples. Including a "trending" regime in the model captures the effect of representativeness by allowing investors to put more weight on trends than they should.

Conservatism suggests that people may put too little weight on the latest piece of earnings news relative to their prior beliefs. In other words, when they get a good piece of earnings news, they effectively act as if part of the shock will be reversed in the next period, in other words, as if they believe in a "mean-reverting" regime.

For a wide range of parameter values, this model generates post-earnings announcement drift, momentum, long-term reversals and cross-sectional forecasting power for scaled-price ratios. After a single earnings announcement, conservatism dominates, and the investor underreacts to the news. In effect, he believes that part of the earnings shock will be reversed. Since earnings actually follow a random walk, the investor will, on average, be positively surprised by the next announcement, generating post-earnings announcement drift and momentum. After a sequence of good earnings, though, the investor will not only correct his initial conservatism, but also swing too far in the other direction. Perceiving a continued trend in earnings, he will push prices up relative to current earnings. Since actual earnings follow a random walk, the investor will on average be disappointed, generating long-term reversals in returns and a scaled-price ratio effect.

Daniel, Hirshleifer, and Subrahmanyam (1998), DHS henceforth, stress biases in the interpretation of *private*, rather than public information. Imagine that the investor does some research on his own to try to determine a firm's future cashflows. DHS assume that he is overconfident about this information; in particular, they argue that investors are more likely to be overconfident about private information they have worked hard to generate than about public information. If the private information is positive, overconfidence means that investors will push prices up too far relative to fundamentals. Future public information will slowly pull prices back to their correct value, thus generating long-term reversals and a scaled-price effect. To get momentum and a post-earnings announcement effect, DHS assume that the public information alters the investor's confidence in his original private information in an asymmetric fashion, a phenomenon known as self-attribution bias: public news which confirms the investor's research strongly increases the confidence he has in that research. Disconfirming public news, though, is given less attention, and the investor's confidence in the private information remains unchanged. This asymetric response means that initial overconfidence is on average followed by even greater overconfidence, generating momentum.

Chopra, Lakonishok, and Ritter (1992), and La Porta et. al. (1997) provide compelling evidence that supports the idea that investors make irrational forecasts of future cashflows. If, as BSV and DHS argue, long-term reversals and the predictive power of scaled-price ratios are driven by excessive optimism or pessimism about future cashflows followed by a correction, then most of the correction should occur at those times when investors find out that their initial beliefs were too extreme, in other words, at earnings announcement dates. The data strongly confirms this prediction. CLR show that De Bondt and Thaler's "winner" portfolio performs particularly poorly in the few days that surround earnings announcements after portfolio formation. LLSV obtain the same finding for a portfolio of "growth" stocks. It is very hard to give a rational reason for why these portfolios earn such low average returns over just a few days of the year.

Momentum and reversals may also be due to positive feedback trading, when one group of investors buys more of an asset which has recently gone up in value. De Long et al (1990b) present one model of this. The simplest justification for such behavior is extrapolative expectations, where investors' expectations of future returns are based on past returns. This in turn, may be due to representativeness and to the law of small numbers in particular. The same argument made by BSV as to why investors might extrapolate past cashflows too far into the future can be applied here to explain why they might extrapolate past returns too far into the future. DSSW also note that institutional features such as portfolio insurance or margin calls can also generate positive feedback trading.

Positive feedback trading also plays a central role in the model of Hong and Stein (1999), although in this case it emerges endogeneously from more primitive assumptions. In this model, two boundedly rational groups of investors interact. In their context, bounded rationality means that investors are only able to process a subset of available information. "Newswatchers" make forecasts based on private information, but do not condition on past prices. "Momentum traders" condition only on the most recent price change.

Hong and Stein also assume that private information diffuses slowly through the population of newswatchers. Since these investors are unable to extract others' private information from prices, the slow diffusion generates momentum. Momentum traders are then added to the mix. Given what they are allowed to condition on, their optimal strategy is to engage in positive feedback trading: a price increase last period is a sign that good private information is diffusing through the economy. By buying, momentum traders hope to profit from the continued diffusion of information. This behavior preserves momentum, but also generates price reversals: since momentum traders cannot observe the extent of news diffusion, they keep buying even after price has reached fundamental value, generating an overreaction that is only later reversed.

These four models differ most in their explanation of momentum. In two of the models – BSV and HS – momentum is due to an initial underreaction followed by a correction. In DSSW and DHS, it is due to an initial overreaction followed by even more overreaction. Within each pair, the stories are different again.

Hong, Lim, and Stein (1999) present supportive evidence for the view of HS that momentum is due simply to slow diffusion of private information through the economy. They argue that the diffusion of information will be particularly slow among small firms and among firms with low analyst coverage, and that the momentum effect should therefore be more prominent there, a prediction they confirm in the data. They also find that among firms with low analyst coverage, momentum is almost entirely driven by prior losers continuing to lose. They argue that this too, is consistent with a diffusion story. If a firm not covered by analysts is sitting on good news, it will do its best to convey the news to as many people as possible, and as quickly as possible; bad news, however, will be swept under the carpet, making its diffusion much slower.

5.2 Belief-based Models with Institutional Frictions

Some papers have argued that the interaction of investor beliefs and institutional factors may be a fruitful way of thinking about some of the anomalous cross-sectional evidence.

A large class of investors, mutual funds, are not allowed to short stocks. Miller (1977) shows that short sales constraints, when combined with mild assumptions about investor beliefs can themselves generate deviations from fundamental value and in particular, explain why stocks with high price-earnings ratios earn lower returns on average.

Suppose that investors do research on a company and that the private information they gather leads them to hold different opinions. Suppose also that they continue to disagree even when they learn others' private information, an assumption that can be thought of as a form of overconfidence. Those investors with bullish opinions will, of course, hold long positions in the stock; bearish investors want to short the stock, but being unable to do so, they sit out of the market in that stock. Stock prices therefore reflect only the opinions of the most optimistic investors. In particular, stocks which investors disagree about more will have greater optimism built into them, and will therefore have high price-earnings ratios. Since these price earnings ratios are *too* high, subsequent returns will be low.

It is interesting to compare this setup with that of DHS. In the latter paper, stocks only become overpriced when a large group of investors simultaneously unearths particularly good private information. Miller's point is that in the presence of short sales constraints, one does not need systematic bullishness or bearishness – differences in opinion are enough.

Scherbina (2000) performs a direct test of the idea that stocks for which there is greater disagreement will earn lower average returns. Using IBES data on analyst forecasts, she groups stocks into quintiles based on the level of dispersion in analysts' forecasts of current year earnings. She confirms that the highest dispersion portfolio earns lower average returns than the lowest dispersion portfolio.

Chen, Hong, and Stein (2000) also test Miller's idea using "breadth" of ownership – defined roughly as the fraction of mutual funds that hold a particular stock – as a proxy for divergence of opinion about the stock. The more dispersion in opinions there is, the more mutual funds will need to sit out the market due to short sales constraints, leading to lower breadth. CHS predict, and confirm in the data, that stocks experiencing a decrease in breadth subsequently have lower average returns compared to stocks whose breadth increases.

Hong and Stein (1999) analyze the implications of short sales constraints and differences of opinion for higher order moments, and show that they lead to skewness. The intuition is that when a stock's price goes down, more information is revealed: by seeing at what point they enter into the market, we learn the valuations of those investors whose pessimistic views could not initially be reflected in the stock price, because of short sales constraints. When the stock market goes up, the sidelined investors stay out of the market and there is less information revelation. This increase in volatility after a downturn is the source of the skewness.

One prediction of this idea is that stocks which investors disagree about more should exhibit greater skewness. Chen, Hong, and Stein (1999) test this idea using increases in turnover as a sign of investor disagreement. They show that stocks whose turnover increases subsequently display greater skewness.

5.3 Preferences

Earlier, we discussed the paper by Barberis, Huang, and Santos (2001) which tried to explain *aggregate* stock market behavior by combining prospect theory, narrow framing, and a dynamic model of loss aversion. Barberis and Huang (2000) show that applying the same ideas to individual stocks can generate the evidence on long-term reversals and on scaled-price ratios. The key idea is that when investors hold a number of different stocks, narrow framing may induce them to derive utility from gains and losses in the value of *individual* stocks. The specification of this additional source of utility is exactly the same as in BHS, except that it is now applied at the individual stock level instead of at the portfolio level: the investor is loss averse over individual stock fluctuations and the pain of a loss on a specific stock depends on that stock's past performance.

To see how this model generates a value premium, consider a stock which has had poor returns several periods in a row. Precisely because the investor focuses on individual stock gains and losses, he finds this very painful and becomes especially sensitive to the possibility of further losses on the stock. In effect, he perceives the stock as riskier, and discounts its future cashflows at a higher rate: this lowers its price-earnings ratio and leads to higher subsequent returns, generating a value premium. In one sense, this model is narrower than those in the "beliefs" section, as it does not claim to address momentum. In another sense, it is broader, in that it simultaneously explains the equity premium and derives the riskfree rate endogeneously.

The models we have described in Sections 5.1, 5.2, and 5.3 have focused primarily on momentum, long-term reversals, the predictive power of scaled-price ratios and post-earnings announcement drift. What about the other examples of anomalous evidence with which we began Section 5? In Section 7, we will argue that the long-run return patterns following equity issuance and repurchases may be the result of rational managers responding to the kinds of noise traders analyzed in the preceding behavioral models. In short, if investors cause prices to swing away from fundamental value, managers may try to time these cycles, issuing equity when it is overpriced, and repurchasing it when it is cheap. In such a world, equity issues will indeed be followed by low returns, and repurchases by high returns. The models we have discussed so far do not, however, shed light on the size anomaly, nor on the dividend announcement event study.

6 Application: Closed-end Funds and Comovement

6.1 Closed-end Funds

Closed-end funds differ from more familiar open-end funds in that they only issue a fixed number of shares. These shares are then traded on exchanges: an investor who wants to buy a share of a closed-end fund must go to the exchange and buy it from another investor at the prevailing price. By contrast, should he want to buy a share of an open-end fund, the fund would create a new share and sell it to him at the fund's net asset value, or NAV, the per share market value of its asset holdings.

The central puzzle about closed-end funds is that fund share prices differ from NAV. The typical fund trades at a discount to NAV of about 10% on average, although the difference between price and NAV varies substantially over time. When closed-end funds are created, the share price is typically above NAV; when they are terminated, either through liquidation or open-ending, the gap between price and NAV closes.

A number of rational explanations for the average closed-end fund discount have been proposed. These include expenses, expectations about future fund manager performance, and tax liabilities. These factors can go some way to explaining certain aspects of the closed-end fund puzzle. However, none of them can satisfactorily explain *all* aspects of the evidence. For example, it is possible to use agency costs explain why funds usually sell at discounts, but agency costs cannot explain why funds sometimes sell at substantial premia (unless agency costs can be negative) nor why discounts tend to vary from week to week and co-vary with each other.

Lee, Shleifer, and Thaler (1991), LST henceforth, propose a simple behavioral view of these closed-end fund puzzles. They argue that some of the individual investors who are the primary owners of closed-end funds are noise traders, exhibiting irrational swings in their expectations about future fund returns. Sometimes they are too optimistic, while at other times, they are too pessimistic. Changes in their sentiment affect share prices and hence also the difference between price and net asset value.[8]

This view provides a clean explanation of all aspects of the closed-end fund puzzle. Owners of closed-end funds have to contend with two sources of risk: fluctuations in the value of the funds' assets, and fluctuations in noise trader sentiment. If this second risk is systematic – we return to this issue shortly – rational investors will demand compensation for it. In other words, they will require that the fund's shares trade at a discount to NAV.

[8]For the noise traders to affect the *difference* between price and NAV rather than just price, it must be that they are more active traders of closed-end fund shares than they are of assets owned by the funds. As evidence for this, LST point out that while funds are primarily owned by individual investors, the funds' assets are not.

This also explains why it is possible to sell new closed-end funds at a premium. Entrepreneurs will choose to create closed-end funds at times of investor exuberance, when they know that they can sell investors fund shares for more than they are worth. On the other hand, when a closed-end fund is liquidated, investors no longer have to worry about changes in noise trader sentiment because they know that at liquidation, the fund price must equal NAV. They therefore no longer demand compensation for this risk, and the fund price rises towards NAV.

An immediate prediction of the LST view is that prices of closed-end funds should comove strongly, even if the cashflow fundamentals of the assets held by the funds do not: if noise traders become irrationally pessimistic, they will sell closed-end funds across the board, depressing their prices regardless of cashflow news. LST confirm in the data that closed-end fund discounts are highly correlated.

The LST story depends on noise trader risk being systematic. There is good reason to think that it is. If the noise traders who hold closed-end funds also hold other assets, then positive changes in sentiment, say, will drive down the prices of closed-end funds *and* of their other holdings, making the noise trader risk systematic. To check this, LST compute the correlation of closed-end fund discounts with another group of assets primarily owned by individuals, small stocks. Consistent with the noise trader risk being systematic, they find a strong positive correlation.

6.2 Comovement

The LST model illustrates that behavioral models can make interesting predictions not only about the *average* level of returns, but also about patterns of comovement. In particular, it explains why the prices of closed-end funds comove so strongly, and also why closed-end funds as a class comove with small stocks. This raises the hope that behavioral models might be able to explain other puzzling instances of comovement.

Before studying this in more detail, is it worth setting out the efficient markets view of return comovement. The simplest rational view of return comovement is that it is due to cashflow comovement: there will be a common factor in the returns of a group of assets if there is a common factor in news about their future earnings. There is little doubt that many instances of return comovement can be explained by cashflows: stocks in the automotive industry move together primarily because their earnings are correlated.

The closed-end fund evidence shows that cashflow view of comovement is at best, incomplete: in that case, the prices of closed-end funds comove even though their fundamentals do not.[9] Other evidence is just as puzzling. Froot and Dabora (1999) study Siamese-twin

[9]Bodurtha et. al. (1993) and Hardouvelis et. al. (1994) provide further interesting examples of a delinking between cashflow comovement and return comovement in the closed-end fund market. They find

stocks, which are claims to the same cashflow stream, but are traded in different locations. The Royal Dutch/Shell pair, discussed in Section 2, is perhaps the best known example. If return comovement is simply a reflection of cashflow comovement, these two stocks should be perfectly correlated. In fact, as Froot and Dabora show, Royal Dutch comoves strongly with the S&P 500 index of U.S. stocks, while Shell comoves with the FTSE index of U.K. stocks.

Fama and French (1993) uncover salient common factors in the returns of small stocks, as well as in the returns on value stocks. In order to test the rational view of comovement, Fama and French (1995) investigate whether these strong common factors can be traced to common factors in the earnings of these stocks. While they do uncover a common factor in the earnings of small stocks, as well as in the earnings of value stocks, these cashflow factors are weaker than the factors in returns and there is little evidence that the return factors are driven by the cashflow factors. Once again, there appears to be comovement in returns that has little to do with cashflow comovement.

In response to this evidence, researchers have begun to posit behavioral theories of comovement. LST is one such theory. To state their argument more generally, they start from the observation that many investors choose to trade only a subset of all available securities. As these investors' risk aversion or sentiment changes, they alter their exposure to the particular securities they hold, thereby inducing a common factor in the returns of these securities. Put differently, this view of comovement predicts that there will be a common factor in the returns of securities that are the primary holdings of a specific subset of investors, such as individual investors. We refer to this as the "habitat" view of comovement, since it relies on certain investors having a preferred habitat. This story seems particularly appropriate to thinking about closed-end funds, and also for Froot and Dabora's evidence.

A second behavioral view of comovement was recently proposed by Barberis and Shleifer (2000). They argue that to simplify the portfolio allocation process, many investors first group stocks into categories such as small-cap stocks or automotive industry stocks, and then allocate funds across these various categories. If these categories are also adopted by noise traders, then as these traders move funds from one category to another, the price pressure from their coordinated demand will induce common factors in the returns of stocks that happen to be classified into the same category, even if those stocks' cashflows are largely uncorrelated. In particular, this view predicts that when an asset is added to a category, it should begin to comove more with that category than before.

Barberis, Shleifer, and Wurgler (2001) test this "category" view of comovement by taking a sample of stocks that has been added to the S&P 500, and computing the covariance of these stocks with the S&P 500 both before and after they are included. Based on both univariate and multivariate regressions, they show that upon inclusion, a stock's beta with

that closed-end funds invested in German equities but traded in the U.S. typically comove more with the U.S. stock market than with the German stock market.

the S&P 500 rises significantly, as does the fraction of its variance that is explained by the S&P 500, while its beta with stocks outside the index falls. This result does not sit well with the cashflow view of comovement – addition to the S&P 500 carries no information about the covariance of a stock's cashflows with other stocks' cashflows – but follows naturally in an economy where prices are affected by category-level demand shocks.

7 Application: Investor Behavior

Behavioral finance has also had some success in explaining how certain groups of investors behave, and in particular, what kinds of portfolios they choose to hold and how they trade over time. The goal here is less controversial than in the previous three sections: it is simply to explain the actions of certain investors, whether or not these actions also affect prices. Two factors make this type of research of growing importance for the future. First, as the costs of entering the market have fallen, more and more individual investors are becoming direct investors in the stock market. Second, the world-wide trend toward defined contribution retirement savings plans, and the possibility of individual accounts in social security systems means that individuals are more responsible for their own financial well being in retirement. Thus it is natural to ask how well they are doing at this job.

We now describe some of the evidence on the actions of investors and the behavioral ideas that have been used to explain it.

Insufficient diversification

A large body of evidence suggests that investors diversify their portfolio holdings much less than is recommended by normative models of portfolio choice.

First, investors exhibit a pronounced "home bias". French and Poterba (1991) report that investors in the U.S., Japan and the U.K. allocate 93%, 98%, and 82% of their overall equity investment, respectively, to *domestic* equities. This fact has resisted all attempts at rational explanation. Indeed, normative portfolio choice models that take human capital into account typically advise investors to *short* their national stock market, because of its high correlation with their human capital.

At least two studies have found an analog to home bias *within* countries. Using an especially detailed Finnish data set, Grinblatt and Keloharju (1999) find that investors in that country are much more likely to hold and trade stocks of Finnish firms which are located close to them geographically, which use their native tongue in company reports, and whose chief executive shares their cultural background. Huberman (1999) studies the geographic distribution of shareholders of U.S. Regional Bell Operating Companies (RBOC's) and finds that investors are much more likely to hold shares in their local RBOC than in out-of-state RBOC's. Finally, studies of allocation decisions in 401(k) plans find a strong bias toward holding own company stock: over 30% of defined contribution plan assets in large U.S. companies are invested in employer stock (Benartzi 2001).

In Section 3, we discussed evidence showing that people dislike ambiguous situations, where they feel unable to specify a gamble's probability distribution. These are often situations where the investor lacks information that could be known. On the other hand, people like familiar situations, where they feel they are in a better position than others to evaluate a gamble.

Ambiguity and familiarity offer a simple way of understanding the different examples of insufficient diversification. Investors may find their national stock markets to be more familiar – or less ambiguous – than foreign stock indices; they may find firms situated close to them geographically to be more familiar than those located further away; and they may find their employer's stock more familiar than other stocks.[10] Since familiar stocks are attractive, people will invest heavily in those, and invest little or nothing at all in ambiguous stocks. Their portfolios will therefore appear undiversified relative to the predictions of standard models that ignore the investor's degree of confidence in the probability distribution of a gamble.

Naive Diversification

Benartzi and Thaler (2001) find that when people *do* diversify, they do so in a naive fashion. In particular, they provide evidence that in 401(k) plans, many people seem to use the simple strategy of allocating $1/n$ of their savings to each of the n available investment options, whatever those options are. Some evidence that people think in this way comes from the laboratory. Benartzi and Thaler ask subjects to make an allocation decision in each of the following three conditions: first, between a stock fund and a bond fund; next, between a stock fund and a balanced fund, which invests 50% in stocks and 50% in bonds; and finally, between a bond fund and a balanced fund. They find that in all three cases, a 50:50 split across the two funds is a popular choice, although of course this leads to very different effective choices between stocks and bonds. The average allocation to stocks in the three conditions was 54%, 73%, and 35% respectively.

The $1/n$ diversification heuristic predicts that in 401(k) plans which predominantly offer stock funds, investors will end up allocating more to stocks. Benartzi and Thaler test this in a sample of 170 large retirement saving plans. They divide the plans into three groups based on the fraction of funds – low, medium, or high – they offer that are stock funds. The allocation to stocks increases across the three groups, from 36% to 65% to 85%, confirming the initial prediction.

Excessive Trading

[10] Particularly relevant to this last point is survey data indicating that people consider their own company stock to be less risky than a diversified index.

One of the clearest predictions of a rational model of investing is that there is very little trading. In a world where rationality is common knowledge, I am reluctant to buy if you are ready to sell. In contrast to this prediction, the volume of trading on the world's stock exchanges is very high. Furthermore, studies of individuals and institutions suggest that both groups trade more than can be justified on rational grounds.

Barber and Odean (2000) study trading activity in a large sample of accounts obtained from a national discount brokerage firm. They find that after carefully taking trading costs into account, the average return of investors in their sample is well below the return of standard benchmarks. Put simply, these investors would do a lot better if they traded less. The underperformance is partly due to transaction costs, but also to poor security selection: in a similar data set, Odean (1999) finds that the average gross return of stocks that investors in his sample buy, over the year after they buy them, is considerably lower than the average gross return of stocks that they sell, over the year after they sell them.

The most prominent behavioral explanation of such excessive trading is overconfidence: people believe that they have information strong enough to justify a trade, while in fact the information is too weak to warrant any action. Given Odean's (1999) evidence, the situation may be even worse: not only do people think that they have information when they don't, but they may even misinterpret valid information.

The overconfidence hypothesis predicts that people who are more overconfident will trade more and, because of transaction costs, earn lower returns. There is further evidence consistent with this. Barber and Odean (2000) show that the investors in their sample who trade the most earn by far the lowest average returns. Building on evidence than men are more overconfident than women, Barber and Odean (2001) predict and confirm that men trade more and earn lower returns on average.

The Selling Decision

Several studies find that investors are reluctant to sell assets that are trading at a loss relative to the price at which they were purchased, a phenomenon labelled the "disposition effect" by Shefrin and Statman (1985). Odean (1998) finds, for example, that the individual investors in his sample are more likely to sell stocks which have gone up in value relative to their purchase price, rather than stocks which have lost value.

It is hard to explain this behavior on rational grounds. Tax considerations point to the selling of losers, not winners. Nor can one argue that investors rationally sell the winners because of information that their future performance will be poor. Odean reports that the average performance of stocks that people sell is better than that of stocks they hold on to.

Odean suggests two behavioral explanations. First, investors may have an irrational belief in mean-reversion. A second possibility relies on prospect theory and narrow framing. We have used these ingredients before, but this time it is not loss aversion that is central, but rather the concavity (convexity) of the value function in the region of gains (losses).

To see the argument, suppose that a stock that was originally bought at \$50 now sells for \$55. Should the investor sell it at this point? Suppose that the gains and losses of prospect theory refer to the sale price minus the purchase price. In that case, the utility from selling the stock now is $v(5)$. Alternatively, the investor can wait another period, whereupon we suppose that the stock could go to \$50 or \$60 with equal probability; in other words, we abstract from belief-based trading motives by saying that the investor expects the stock price to stay flat. The expected value of waiting and selling next period is then $\frac{1}{2}v(0) + \frac{1}{2}v(10)$. Since the value function v is concave in the region of gains, the investor sells now. In a different scenario, the stock may currently be trading at \$45. This time, the comparison is between $v(-5)$ and $\frac{1}{2}v(-10) + \frac{1}{2}v(0)$, assuming a second period distribution of \$40 and \$50 with equal probability. Convexity of v pushes the investor to wait. Intuitively, as long as the stock is still in the portfolio, the investor can tell himself that it may recover, but if the stock is actually sold, he has to admit that he made a mistake.

The disposition effect is not confined to individual stocks. In an innovative study, Genesove and Mayer (2001) find evidence of a reluctance to sell at a loss in the housing market. They show that sellers whose expected selling price is below their original purchase price set an asking price that exceeds the asking price of other sellers by about 30% of the difference between the previous selling price and current market value. Moreover, this is not simply wishful thinking on the sellers' part that is later corrected by the market. Sellers facing a possible loss do actually transact at considerably higher prices than other sellers.

The Buying Decision

Odean (1999) also presents useful information about the stocks investors in his sample choose to buy. Unlike "sells", which are mainly prior winners, "buys" are evenly split between prior winners and losers. Conditioning on the stock being a prior winner (loser) though, the stock is a big winner (loser). In other words, a good deal of the action is in the extremes.

Odean argues that the results for stocks are in part due to an attention effect. When buying a stock, people do not tend to systematically sift through the thousands of listed shares until they find a good "buy." They typically buy a stock that has caught their attention and perhaps the best attention draw is extreme past performance, whether good or bad.

Another possibility is that some investors systematically seek out and buy stocks with very good past performance because they believe that this good performance will continue; a naive use of the representativeness heuristic would lead to such beliefs. Other investors may seek out stocks with extremely poor prior performance because they believe that such stocks are undervalued and will rebound.

Our discussion of the buying decision focuses on belief-based stories, while stories about preferences figure more prominently among explanations for selling behavior and for the disposition effect in particular. This may seem needlessly messy, but it is worth noting that in some ways the buying and selling decisions are very different and therefore may indeed warrant different theories. Since people are reluctant to sell short, they typically sell stocks that they already own. When buying stocks, people have a huge range to choose from, and different factors may enter the decision.

8 Application: Corporate Finance

8.1 Security Issuance, Capital Structure, and Investment

A number of recent papers have argued that many empirical facts about the timing of firms' security issuance, their capital structure, and investment patterns may be the result of actions that rational managers take when faced with irrational investors like those described in earlier sections.

In order to make this case, we must first analyze what a rational manager should do if he perceives his firm's securities to be mispriced. Stein (1996) provides a useful framework for thinking about this. To a first approximation, his analysis can be summarized as follows. Suppose that a manager thinks his firm's stock price is overvalued and that he is interested in maximizing the firm's true value, or in other words, the stock price that will prevail once mispricing has worked its way out of valuations. The first, and more obvious action he should take is to issue more shares so as to take advantage of investor exuberance.

More subtly, though, Stein shows that he should *not* channel the fresh capital into any actual new investment, but instead keep it as cash, or invest it in other fairly priced capital market securities. While investors' exuberance means that, in *their* view, the firm has many positive net present value (NPV) projects it could undertake, the rational manager knows that these projects are not in fact, positive NPV, and that in the interest of true firm value, should be avoided. Conversely, if the manager thinks that his firm's stock price is irrationally low, he should repurchase shares at the advantageously low price but not scale back actual investment. In short, irrational investors may affect the timing of security issuance, but they should not affect the firm's investment plans.[11] We refer to this view of the world as the

[11] Some caveats to this last point about investment should be noted. If a rational manager thinks investors are excessively optimistic, he may still increase investment even if he knows that the new projects have a negative NPV. For example, if investors see him refusing to undertake projects they perceive as profitable, they may try to have him fired. Alternatively, if the manager wants to keep issuing overvalued equity for a period of time, actually engaging in some new investment may be a good way of stoking the fire and prolonging investor frenzy. Finally, just because the manager is rational does not mean he will choose to maximize the firm's true value. The agency literature has argued that managers may maximize other objectives – the size of their firm, say – as way a way of enhancing their prestige. Such a manager might use

"market timing" model.

Interestingly, the evidence on both security issuance and investment is quite consistent with this framework. First, at the aggregate level, the share of new equity issues among total new issues – the "equity share" – is higher when the overall stock market is more highly valued. In fact, Baker and Wurgler (2000a) show that the equity share is a reliable predictor of future stock returns: a high share predicts low, and sometimes negative stock returns. This is consistent with managers timing the market and issuing more equity at market peaks, just before it sinks back to more realistic valuation levels.

At the individual firm level, a number of papers have shown that the book-to-market ratio of a firm is a good cross-sectional predictor of new equity issuance (see Koracjzyk, Lucas, Macdonald (1991), Jung, Kim, Stulz (1996), Loughran, Ritter, Ridquist (1994), Pagano, Panneta, Zingales (1998), Baker and Wurgler 2000b). Firms with high valuations issue more equity while those with low valuations repurchase their shares. Moreover, long term stock returns after an IPO or SEO are very low (Loughran, Ritter 1995), while long term returns after the announcement of a repurchase are high (Ikenberry, Lakonishok, Ritter, 1996). Once again, this evidence is consistent with managers timing the market in their own securities.

The success of the market timing framework in predicting patterns of security issuance offers the hope that it might also be the basis of a successful theory of capital structure. After all, a firm's capital structure simply represents its cumulative financing decisions over time. Consider, for example, two firms which are similar in terms of characteristics like firm size, profitability, fraction of tangible assets, and current market-to-book ratio, which have traditionally been thought to affect capital structure. Suppose, however, that in the past, the market-to-book ratio of firm A has reached much higher levels than that of firm B. Since, under the market timing theory, managers of firm A may have issued more shares at that time to take advantage of possible overvaluation, firm A may have more equity in its capital structure today. In a remarkable recent paper, Baker and Wurgler (2000b) confirm this prediction. In particular, they show that all else equal, a firm's maximum historical market-to-book ratio is a very good cross-sectional predictor of the fraction of equity in the firm's capital structure today.[12]

Still more support for the market timing view of capital structure comes from survey evidence. Graham and Harvey (2000) report that 67% of surveyed CFO's said that "the amount by which our stock is undervalued or overvalued" was an important consideration when issuing common stock.

All the above facts are encouraging for proponents of the market timing story. Never-

---

investor exuberance as a cover for doing some negative NPV "empire building" projects of his own.

[12]Baker and Wurgler (2000b) get even stronger results using the average historical market-to-book as an explanatory variable, where the average is taken across time periods when the firm actually raised external financing – whether debt or equity.

theless, to really clinch the argument, the evidence on investment behavior is crucial. This is because much of the evidence so far can be explained with a different story. Suppose that investors do cause swings in firms' stock prices, but that the manager does not perceive these swings as irrational. Moreover, suppose that the manager follows a pecking-order rule for capital structure, perhaps because of asymmetries of information between investors and managers which make external financing somewhat costly.

In this case, if the firm's stock price goes up, the manager agrees with investors that he now has many more attractive new projects to potentially undertake. If other forms of financing have been exhausted, he will need to raise equity to finance these new projects. Therefore, in the same way as with the market timing hypothesis, high valuations will be accompanied by more equity issuance. A crucial difference between the two stories, though, is that in this case, the manager will actually use the new funds to do more investment. Broadly speaking, stock prices and investment will move together more than they do under the market timing theory.

In general, the evidence on investment is more supportive of market timing stories although it is by no means conclusive. In aggregate data, Blanchard, Rhee, and Summers (1992) find that movements in price unrelated to movements in fundamentals have only weak forecasting power for future investment: the effects are marginally statistically significant and weak in economic terms. To pick out two particular historical examples: the rise in stock prices through the 1920's did not lead to a commensurate rise in investment, nor did the crash of 1987 slow investment down appreciably. Morck, Shleifer, and Vishny (1993) reach similar conclusions using firm level data. In their recent work on capital structure, Baker and Wurgler (2000b) obtain even stronger results. Not only do firms with higher market-to-book ratios in their past have more equity in their capital structure today, but those equity funds are typically used to increase cash balances and *not* to finance new investment.

8.2 Dividends

A major open question in corporate finance asks why firms pay dividends. Historically, dividends have been taxed at a higher rate than capital gains. This means that stockholders who pay taxes would always prefer that the firm repurchases shares rather than paid a dividend. Since the tax exempt shareholders would be indifferent between the dividend payment and the share repurchase, the share repurchase is a Pareto improving action. Why then, do investors seem perfectly happy to accept a substantial part of their return in the form of dividends? Or, using the behavioral language, why do firms choose to frame a part of their earnings as an explicit payment to stockholders, and in so doing, make some of their shareholders worse off?

Shefrin and Statman (1984) propose a number of plausible behavioral explanations for why investors exhibit a preference for dividends. Their first idea relies on the notion of self-control. Many people exhibit self-control problems. On the one hand, we want to deny ourselves an indulgence, but on the other hand, we quickly give in to temptation: today, we tell ourselves that tomorrow we will not overeat, and yet, when tomorrow arrives, we again eat too much. To deal with self-control problems, people often set rules, such as "bank the wife's salary, and only spend from the husband's paycheck". Another very natural rule people might create to prevent themselves from overconsuming their wealth is "only consume the dividend, but don't touch the portfolio capital". In other words, people may like dividends because dividends help them surmount self-control problems through the creation of simple rules.

A second rationale for dividends is based on mental accounting: by designating an explicit dividend payment, firms make it easier for investors to segregate gains from losses and hence to increase their utility. To see this, consider the following example. Over the course of a year, the value of a firm has increased by \$10 per share. The firm could choose *not* to pay a dividend and return this increase in value to investors as a \$10 capital gain. Alternatively, it could pay a \$2 dividend, leaving an \$8 capital gain. Applying prospect theory, investors will code the first option as $v(10)$, and the second as $v(2) + v(8)$ which gives a higher utility, due to the concavity of v in the domain of gains.

This manipulation is equally useful in the case of losses. A firm whose value has declined by \$10 per share over the year can offer investors a \$10 capital loss or a \$12 capital loss combined with a \$2 dividend gain. Once again, the utility of the second, $v(2) + v(-12)$ is greater than the utility of the first, $v(-10)$, this time because of the convexity of v in the domain of losses.

The utility enhancing trick in these examples depends on investors segregating the overall gain or loss into different components. The key insight of Shefrin and Statman is that by paying dividends, firms make it easier for investors to perform this segregation.

Finally, Shefrin and Statman argue that by paying dividends, firms help investors avoid regret. Regret is a frustration that people feel when they imagine having taken an action that would have led to a more desirable outcome. It is stronger for errors of commission – cases where people suffer because of an action they took – than for errors of omission – where people suffer because of an action they *failed* to take.

Consider a company which does not pay a dividend. In order to finance consumption, an investor has to sell stock. If the stock subsequently goes up in value, the investor feels substantial regret because the error is one of commission: he can readily imagine how not selling the stock would have left him better off. If the firm had paid a dividend and the investor was able to finance his consumption out of it, a rise in the stock price would not have caused so much regret. This time, the error would have been one of omission: to be better off, the investor would have had to reinvest the dividend.

Shefrin and Statman try to explain why firms pay dividends at all. Another question asks how dividend paying firms decide on the size of their dividend. The classic paper on this subject is Lintner (1956). His treatment is based on extensive interviews with executives of large American companies in which Linter asked the respondent, often the CFO, how the firm set dividend policy. Based on these interviews Lintner proposed what we would now call a behavioral model. In Linter's model, firms first establish a target dividend payout rate based on notions of fairness, in other words, what portion of the earnings is it fair to return to the shareholders. Then, as earnings increase and the dividend payout ratio falls below the target level, firms increase dividends only when they are confident that they will not have to reduce them in the future.

There are several behavioral aspects to this model. First, the firm is not setting the dividend to maximize firm value or shareholder (after-tax) wealth. Second, perceptions of fairness are used to set the target payout rate. Third, the asymmetry between an increase in dividends and a decrease is explicitly considered. Although fewer firms now decide to start paying dividends, for those that do Lintner's model appears to be valid to this day (see Fama and French, 2001, and Benartzi, Michaely, and Thaler, 1997).

8.3 Models of Managerial Irrationality

The theories we have discussed so far interpret the data as the result of actions taken by rational managers in response to irrationality on the part of investors. Other papers have argued that some aspects of managerial behavior are the result of irrationality on the part of managers themselves.

Much of Section 2 was devoted to thinking about whether rational agents might be able to correct dislocations caused by irrational traders. Analogously, before we consider models of irrational managers, we should ask to what extent rational agents can undo the effects of such managers.

On reflection, it doesn't seem any easier to deal with irrational managers than irrational investors. It is true that many firms have mechanisms in place designed to solve agency problems and to keep the manager's mind focused on maximizing firm value: giving him stock options for example, or saddling him with debt. The problem is that these mechanisms are unlikely to have much of an effect on irrational managers. These managers *think* that they are maximizing firm value, even if in reality, they are not. Since they think that they are already doing things right, stock options or debt are unlikely to change their behavior.

In the best known paper on managerial irrationality, Roll (1986) argues that much of the evidence on takeover activity is consistent with an economy in which there are *no* overall gains to takeovers, but in which managers are overconfident, a theory he terms the "hubris hypothesis". When managers think about taking over another firm, they conduct a valuation analysis of that firm, taking synergies into account. If managers are overconfident about the accuracy of their analysis, they will be too quick to launch a bid when their valuation exceeds the market price of the target. Just as overconfidence among individual investors may lead to excessive trading, so overconfidence among managers may lead to excessive takeover activity.

The main predictions of the hubris hypothesis are that there will be a large amount of takeover activity, but that the total combined gain to bidder and target will be zero; that on the announcement of a bid, the price of the target will rise and the value of the bidder will fall by a similar amount. Roll examines the available evidence and concludes that it is impossible to reject any of these predictions.

Heaton (1997) analyses the consequences of managerial optimism in which managers overestimate the probability that the future performance of their firm will be good. He shows that it can explain pecking order rules for capital structure: since managers are optimistic relative to the capital markets, they believe their equity is undervalued, and are therefore reluctant to issue it unless they have exhausted internally generated funds or the debt market. Managerial optimism can also explain the correlation of investment and cashflows: when cashflow is low, managers reluctance to use external markets for financing means that they will forgo an unusually large number of projects, lowering investment at the same time. Finally, free cash flow is especially dangerous in this world, as managers may use it to invest in projects they perceive as having a positive NPV even though they do not.

9 Conclusion

Behavioral finance is a young field, with its formal beginnings in the 1980s. Much of the research we have discussed was completed in the past five years. Where do we stand? Substantial progress has been made on numerous fronts.

Empirical investigation of apparently anomalous facts. When De Bondt and Thaler's (1985) paper was published, many scholars thought that the best explanation for their findings was a programming error. Since then their results have been replicated numerous times by authors both sympathetic to their view and by those with alternative views. At this stage, we think that most of the empirical facts are agreed upon by most of the profession, although the interpretation of those facts is still in dispute. This is progress. If we all agree that the planets do orbit the sun, we can focus on understanding why.

Limits of Arbitrage. Twenty years ago, many financial economists thought that the Efficient Markets Hypothesis had to be true because of the forces of arbitrage. We now understand that this was a naive view, and that the limits of arbitrage can permit substantial mispricing. It is now also understood by most that the absence of a profitable investment strategy, because of risks and costs, does not imply the absence of mispricing. Prices can be very wrong without creating profit opportunities.

Understanding Bounded Rationality. Thanks largely to the work of cognitive psychologists such as Daniel Kahneman and Amos Tversky, we now have a long list of robust empirical findings that catalogue some of the ways in which actual humans form expectations and make choices. There has also been progress in writing down formal models of these processes, with prospect theory being the most notable. Economists once thought that behavior was either rational or impossible to formalize. We now know that models of bounded rationality are both possible and also much more accurate descriptions of behavior than purely rational models.

Behavioral Finance Theory Building. In the past few years there has been a burst of theoretical work modelling financial markets with less than fully rational agents. These papers relax the assumption of complete rationality either through the belief formation process or through the decision making process. Like the work of psychologists discussed above, these papers are important existence proofs, showing that it is possible to think coherently about asset pricing while incorporating salient aspects of human behavior.

Investor Behavior. We have now begun the important job of trying to document and understand how investors, both amateurs and professionals, make their portfolio choices. Until recently such research was notably absent from the repertoire of financial economists, perhaps because of the mistaken belief that asset pricing can be modeled without knowing anything about the behavior of the agents in the economy.

This is a lot of accomplishment in a short period of time, but we are still much closer to the beginning of the research agenda than we are to the end. We know enough about the perils of forecasting to realize that most of the future progress of the field is unpredictable. Still, we cannot resist venturing a few observations on what may be coming next.

First, much of the work we have summarized is narrow. Models typically capture something about investors' beliefs, or their preferences, or the limits of arbitrage, but not all three. This comment applies to most research in economics, and is a natural implication of the fact that researchers are boundedly rational too. Still, as progress is made we expect theorists to begin to incorporate more than one strand into their models.

An example can, perhaps, illustrate the point. The empirical literature repeatedly finds that the asset pricing anomalies are more pronounced in small and mid-cap stocks than in the large cap sector. It seems likely that this finding reflects limits of arbitrage: the costs of trading smaller stocks are higher, and the low liquidity keeps many potential arbitrageurs uninterested. While this observation may be an obvious one, it has not found its way into formal models. We expect investigation of the interplay between limits of arbitrage and cognitive biases to be an important research area in the coming years.

Second, there are obviously competing behavioral explanations for some of the empirical facts. Some critics view this as a weakness of the field. It is sometimes said that the long list of cognitive biases summarized in Section 3 offer behavioral modelers so many degrees of freedom that anything can be explained. We concede that there are numerous degrees of freedom, but note that rational modelers have just as many options to choose from. As Arrow (1986) has forcefully argued, rationality per se does not yield many predictions. The predictions come from auxiliary assumptions.

There is really only one scientific way to compare alternative theories, behavioral or rational, and that is with empirical tests. One kind of test looks for novel predictions the theory makes. For example, LST (1991) test their model's prediction that small firm returns will be correlated with closed-end fund discounts, while Hong, Lim and Stein (2000) test the implication of the Hong and Stein (1999) model that momentum will be stronger among stocks with thinner analyst coverage.

Another sort of test is to look for evidence that agents actually behave the way a model claims they do. The Odean (1998) and Genesove and Mayer (2000) investigations of the disposition effect using actual market behavior fall into this category. Bloomfield et. al. (2000) offers an experimental test of the behavior theorized by BSV (1998). Of course, such tests are never airtight, but we should be skeptical of theories based on behavior that is undocumented empirically. Since behavioral theories claim to be grounded in realistic assumptions about behavior, we hope behavioral finance researchers will continue to give their assumptions empirical scrutiny. We would urge the same upon authors of rational theories.[13]

We have two predictions about the outcome of the exercise of direct tests of the assumptions of economic models. First, we will find out that most of our current theories, both rational and behavioral, are wrong. Second, we will produce better theories.

[13]Directly testing the validity of a model's assumptions is not common practice in economics, perhaps because of Milton Friedman's influential argument that one should evaluate theories based on the validity of their predictions rather than the validity of their assumptions. Whether or not this is sound scientific practice, we note that much of the debate over the past 20 years has occured precisely because the evidence has not been consistent with the theories, so it may be a good time to start worrying about the assumptions. If a theorist wants to claim that fact X can be explained by behavior Y, it seems prudent to check whether people actually do Y.

10 References

Anderson E., Lars Hansen, and Thomas Sargent (1999), "Risk and Robustness in Equilibrium," Working Paper, University of Chicago.

Baker, Malcolm, and Jeffrey Wurgler (2000), "The Equity Share in New Issues and Aggregate Stock Returns," *Journal of Finance* 55, 2219-2257.

Baker, Malcolm, and Jeffrey Wurgler (2000), "Market Timing and Capital Structure", working paper, Yale University.

Banz, Rolf (1981), "The Relation between Return and Market Value of Common Stocks", *Journal of Financial Economics* 9, 3-18.

Barber, Brad, and John Lyon (1997), "Detecting Long-run Abnormal Stock Returns: the Empirical Power and Specification of Test Statistics," *Journal of Financial Economics* 43, 341-372.

Barber, Brad, and Terrance Odean (2000), "Online Investors: Do the Slow Die First?", Working paper, UC-Davis.

Barber, Brad, and Terrance Odean (2001), "Boys will be Boys: Gender, Overconfidence, and Common Stock Investment," forthcoming *Quarterly Journal of Economics.*

Barberis, Nicholas, and Ming Huang (2001), "Mental Accounting, Loss Aversion, and Individual Stock Returns", forthcoming, *Journal of Finance*

Barberis, Nicholas, Ming Huang, and Tano Santos (2001), "Prospect Theory and Asset Prices", forthcoming, *Quarterly Journal of Economics.*

Barberis, Nicholas, and Andrei Shleifer (2000), "Style Investing," Working Paper, University of Chicago.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny (1998), "A Model of Investor Sentiment", *Journal of Financial Economics* 49, 307-345.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler (2001), "Why do Stocks Comove? Evidence from S&P 500 Inclusions," Working Paper, University of Chicago.

Barsky, Robert, and Brad De Long (1992), "Why does the stock market fluctuate?", *Quarterly Journal of Economics* 107, 291-311.

Benartzi, Shlomo (2001), "Excessive Extrapolation and the Allocation of 401(k) Accounts to Company Stock," forthcoming, *Journal of Finance.*

Benartzi, Shlomo, Roni Michaely and Richard Thaler (1997), "Do Changes in Dividends

Signal the Future or the Past?" *Journal of Finance* 52, 1007-34.

Benartzi, Shlomo, and Richard Thaler (1995), "Myopic Loss Aversion and the Equity Premium Puzzle", *Quarterly Journal of Economics* 110, 75-92.

Benartzi, Shlomo, and Richard Thaler (1998), "Naïve Diversification Strategies in Defined Contribution Savings Plans", working paper, UCLA.

Benartzi, Shlomo, and Richard Thaler (1999), "Risk Aversion or Myopia? Choices in Repeated Gambles and Retirement Investments," *Management Science* 45, 364-381.

Bernard, Victor (1992), "Stock Price Reactions to Earnings Announcements", in Thaler (ed.) *Advances in Behavioral Finance*, ch.11.

Blanchard, Olivier, Changyong Rhee, and Lawrence Summers (1993), "The Stock Market, Profit, and Investment", *Quarterly Journal of Economics.*

Bloomfield, Robert, and Jeffrey Hales (2000), "Predicting the Next Step of a Random Walk: Experimental Evidence of Regime-Shifting Beliefs," Working Paper, Cornell University.

Bodurtha J., Kim, D., and C.M. Lee (1993) "Closed-end Country Funds and U.S. Market Sentiment," *Review of Financial Studies* 8, 879-918.

Brav, Alon (2000), "Inference in Long-horizon Event Studies," *Journal of Finance* 55, 1979-2016.

Brunnermeier, Markus (2001), *Asset Pricing under Asymmetric Information – Bubbles, Crashes, Technical Analysis, and Herding*, Oxford University Press.

Camerer, Colin (1995), "Individual Decision Making", in Kagel and Roth (eds.), *Handbook of Experimental Economics*, Princeton University Press.

Camerer, Colin, and Martin Weber (1992), "Recent Developments in Modeling Preferences: Uncertainty and Ambiguity," *Journal of Risk and Uncertainty* 5, 325-70.

Campbell, John Y. (1998), "Asset Prices, Consumption, and the Business Cycle", forthcoming in Taylor and Woodford (eds.) *Handbook of Macroeconomics*, North-Holland.

Campbell, John Y., and John Cochrane (1999), "By Force of Habit: A Consumption-based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy* 107, 205-51.

Campbell, John Y. and Robert J. Shiller (Winter 1998), "Valuation Ratios and the Long-Run Stock Market Outlook", *Journal of Portfolio Management.*

Campbell, John Y. and Luis Viceira (2001), *Strategic Asset Allocation*, Oxford University Press.

Chan, Louis, Narasimhan Jegadeesh, and Josef Lakonishok (1996), "Momentum Strategies," *Journal of Finance* 51, 1681-713.

Chen, Joseph, Harrison Hong, and Jeremy Stein (2000), "Breadth of Ownership and Stock Returns", Working paper, Stanford University.

Chen, Joseph, Harrison Hong and Jeremy Stein (2001), "Forecasting Crashes: Trading Volume, Past Returns and Conditional Skewness in Stock Prices", forthcoming *Journal of Financial Economics.*

Chew, S. (1983), "A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox," *Econometrica* 51, 1065-1092.

Chew, S. (1989), "Axiomatic Utility Theories with the Betweenness Property," *Annals of Operations Research* 19, 273-98.

Chew, S., and K. MacCrimmon (1979), "Alpha-nu Choice Theory: An Axiomatization of Expected Utility," Working Paper, University of British Columbia.

Chopra, Navin, Josef Lakonishok, and Jay Ritter (1992), "Measuring Abnormal Performance: Do stocks overreact?", *Journal of Financial Economics* 31: 235-268.

Cochrane, John, "Where is the Market Going? Uncertain Facts and Novel Theories", Economic Perspectives, Federal Reserve Bank of Chicago, November/December 1997.

Cochrane, John, "New Facts in Finance", Economic Perspectives, Federal Reserve Bank of Chicago, Third Quarter 1999.

Daniel, Kent and Sheridan Titman (1997), "Evidence on the Characteristics of Cross-Sectional Variation in Stock Returns", *Journal of Finance* 52, 1-33.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam (1998), "Investor Psychology and Security Market Under- and Overreactions", *Journal of Finance* 53, 1839-1885

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam (2001), "Overconfidence, Arbitrage, and Equilbrium Asset Pricing," *Journal of Finance* 56, 921-965.

De Bondt, Werner, and Richard Thaler (1985), "Does the Stock Market Overreact?", *Journal of Finance* 40, 793-808

De Bondt, Werner, and Richard Thaler (1995), "Financial Decision Making in Markets and Firms", in Jarrow, Maksimovic, and Ziemba (eds.) Finance, Elsevier-North Holland.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert Waldmann (1990a), "Noise Trader Risk in Financial Markets", *Journal of Political Economy* 98, 703-738

De Long, Brad, Andrei Shleifer, Lawrence Summers, Michael Waldmann (1990b), "Positive Feedback Investment Strategies and Destabilizing Rational Speculation", *Journal of Finance* 45, 375-395

Dekel, E. (1986), "An Axiomatic Characterization of Preferences Under Uncertainty: Weakening the Independence Axiom," *Journal of Economic Theory* 40, 304-18.

Edwards, Ward (1986), "Conservatism in Human Information Processing," in B. Kleinmutz (ed.), *Formal Representation of Human Judgment*, New York: John Wiley and Sons.

Ellsberg, D. (1961), "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics* 75, 643-69.

Fama, Eugene F. (1970), "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance* 25, 383-417.

Fama, Eugene F. (1998), "Market Efficiency, Long-Term Returns, and Behavioral Finance", *Journal of Financial Economics* 49, 283-307.

Fama, Eugene (1991), "Efficient Capital Markets: II", *Journal of Finance* 46, 1575-1618.

Fama, Eugene F. and Kenneth R. French (1988), "Dividend Yields and Expected Stock Returns", *Journal of Financial Economics* 22, 3-25.

Fama, Eugene F. and Kenneth R. French (1992), "The Cross-Section of Expected Stock Returns", *Journal of Finance* 47, 427-465.

Fama, Eugene F. and Kenneth R. French (1993), "Common Risk Factors in the Returns of Bonds and Stocks", *Journal of Financial Economics* 33, 3-56.

Fama, Eugene F., and Kenneth R. French (1995), "Size and Book-to-Market Factors in Earnings and Returns," *Journal of Finance* 50, 131-155.

Fama, Eugene F. and Kenneth R. French (1996), "Multifactor Explanations of Asset Pricing Anomalies", *Journal of Finance* 51, 55-84.

Fama, Eugene F. and Kenneth R. French (1998), "Value vs. Growth: The International Evidence," *Journal of Finance* 53, 1975-1999.

Fama, Eugene F., Kenneth R. French and James Davis (2000), "Characteristics, Covariances, and Average Returns 1929-1997," *Journal of Finance* 55, 389-406.

Fama, Eugene F. and Kenneth French (2001), "Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay?" *Journal of Financial Economics* 60, 3-43.

Fox, Craig, and Amos Tversky, "Ambiguity Aversion and Comparative Ignorance," *Quarterly Journal of Economics* 110, 585-603.

French, Kenneth, and James Poterba (1991), "Investor Diversification and International Equity Markets," *American Economic Review* 81, 222-226.

French, Kenneth, William Schwert, and Robert Stambaugh (1987), "Expected Stock Returns and Volatility," *Journal of Financial Economics* 19, 3-30.

Friedman, Milton (1953), "The Case for Flexible Exchange Rates," in *Essays in Positive Economics*, Chicago: University of Chicago Press.

Genesove, and Mayer (2001), "Loss Aversion and Seller Behavior: Evidence from the Housing Market", forthcoming, Quarterly Journal of Economics.

Gilboa Itzhak, and David Schmeidler (1989), "Maxmin Expected Utility with a Non-unique Prior," *Journal of Mathematical Economics* 18, 141-53.

Gneezy, Uri, and Jan Potters (1997), "An Experiment on Risk Taking and Evaluation Periods", *Quarterly Journal of Economics* 112, 631-645.

Graham, John, and Campbell Harvey (2001), "The Theory and Practice of Corporate Finance: Evidence from the Field," *Journal of Financial Economics* 60, 187-243.

Grinblatt, Mark, and Matti Keloharju (1999) "How Distance, Language, and Culture Influence Stockholdings and Trades," *Journal of Finance* 56, 1053-73.

Grinblatt, Mark, and Tobias Moskowitz (1999), "The Cross-section of Expected Returns and its Relation to Past Returns," Working Paper, University of Chicago.

Gul, Faruk (1991), "A Theory of Disappointment in Decision Making under Uncertainty," *Econometrica* 59, 667-86.

Hansen, Lars, and K. Singleton (1983), "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy* 91, 249-268.

Hardouvelis, G., La Porta R., and T. Wizman (1994), "What Moves the Discount on Country Equity Funds?," in Jeffrey Frankel (ed.), *The Internationalization of Equity Markets*, Chicago, The University of Chicago Press.

Harris, Lawrence, and E. Gurel (1986), "Price and Volume Effects associated with Changes in the S&P 500: New Evidence for the Existence of Price Pressure," *Journal of Finance* 41, 851-60.

Heath, Chip, Steven Huddart and Mark Lang (1999), "Psychological Factors and Stock Option Exercises", Quarterly Journal of Economics 114, 601-627.

Heath, Chip, and Amos Tversky (1991), "Preference and Belief: Ambiguity and Competence in Choice under Uncertainty," *Journal of Risk and Uncertainty* 4, 5-28.

Heaton, J.B. (1997), “Managerial Optimism and Corporate Finance”, working paper, University of Chicago.

Hirshleifer, David (2001), “Investor Psychology and Asset Pricing,” *Journal of Finance* 56, 1533-1597.

Hong, Harrison, Terence Lim, and Jeremy Stein (2000), “Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies”, *Journal of Finance* 55, 265-295.

Hong, Harrison, and Jeremy Stein (1999), “A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets”, *Journal of Finance* 54, 2143-2184

Hong, Harrison, and Jeremy Stein (1999), “Differences of Opinion, Rational Arbitrage, and Market Crashes”, working paper, Stanford University.

Huberman, Gur, “Familiarity Breeds Investment”, working paper, Columbia University.

Ikenberry, David, Josef Lakonishok, and Theo Vermaelen (1995), “Market Underreaction to Open Market Share Repurchases”, *Journal of Financial Economics* 39, 181-208.

Jegadeesh, Narasimhan and Sheridan Titman (1993), “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency”, *Journal of Finance* 48, 65-91.

Jung, Kooyul, Yong Cheol Kim, and Rene Stulz, “Timing, Investment Opportunities, Managerial Discretion, and the Security Issue Decision,” *Journal of Financial Economics* 42, 159-185.

Kahneman, Daniel, and Mark Riepe (1998), “Aspects of Investor Psychology”, *Journal of Portfolio Management* 24, 52-65.

Kahneman, Daniel, and Amos Tversky (1974), “Judgment Under Uncertainty: Heuristics and Biases”, *Science* 185, 1124-31.

Kahneman, Daniel, and Amos Tversky (1979), “Prospect Theory: An Analysis of Decision Under Risk”, *Econometrica* 47, 263-91.

Korajczyk, Robert, Deborah Lucas, and Robert MacDonald (1991), “The Effects of Information Releases on the Pricing and Timing of Equity Issues,” *Review of Financial Studies* 4, 685-708.

La Porta, Rafael (1996), “Expectations and the Cross-Section of Returns”, *Journal of Finance* 51, 1715-1742.

La Porta, Rafael, Josef Lakonishok, Andrei Shleifer, and Robert W. Vishny (1994), “Good News for Value Stocks: Further Evidence on Market Efficiency”, *Journal of Finance* 49, 1541-1578.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny (1994), “Contrarian Investment, Extrapolation, and Risk”, *Journal of Finance* 49, 1541-1578.

Lakonishok, Josef and Seymour Smidt (1988), “Are Seasonal Anomalies Real? A Ninety Year Perspective”, *Review of Financial Studies* 3, 257-280.

Lamont, Owen (2000), “Guilty as Charged: Violations of the Law of One Price in Financial Markets,” University of Chicago.

Lamont, Owen, and Richard Thaler (2000), “Can the Market Add and Subtract? Mispricing in Tech Stock Carve-Outs,” Working Paper, University of Chicago.

Lintner, John (1956), “Distribution of Incomes of Corporations among Dividends, Retained Earnings and Taxes”, *American Economic Review* 46, 97-113.

Loughran, Tim, and Jay Ritter (1995), “The New Issues Puzzle”, *Journal of Finance* 50, 23-50.

Loughran, Tim, and Jay Ritter (2000), “Uniformly Least Powerful Tests of Market Efficiency,” *Journal of Financial Economics* 55, 361-389.

Maenhout, Pascal (2000), “Robust Portfolio Rules and Asset Pricing”, working paper, Harvard University.

Markowitz, H. (1952), “The Utility of Wealth,” *Journal of Political Economy* 60, 151-8.

Mehra, Rajnish and Edward Prescott (1985), “The Equity Premium: A Puzzle”, *Journal of Monetary Economics* 15, 145-161.

Miller, Edward (1977), “Risk, Uncertainty, and Divergence of Opinion,” *Journal of Finance* 32, 1151-1168.

Merton, Robert (1987), “A Simple Model of Capital Market Equilibrium with Incomplete Information,” *Journal of Finance* 42 483-510.

Michaely, Roni, Richard Thaler, and Kent Womack, “Price Reactions to Dividend Initiations and Omissions”, *Journal of Finance* 50, 573-608.

Miller, Merton (1986), “Behavioral Rationality in Finance: The Case of Dividends”, in Hogarth and Reder (eds.) *Rational Choice*, University of Chicago Press.

Mitchell, Mark, Todd Pulvino, and Erik Stafford (2000), “Limited Arbitrage in Equity Markets,” forthcoming, *Journal of Finance.*

Mitchell, Mark, and Erik Stafford (2001), “Managerial Decisions and Long-term Stock Price

Performance," *Journal of Business* 73, 287-329.

Modigliani, Franco and Richard Cohn (1974), "Inflation and the Stock Market," *Financial Analysts Journal* 35, 24-44.

Morck, Randall, Andrei Shleifer, and Robert Vishny (1993), "The Stock Market and Investment: Is the Market a Sideshow?" Brookings Papers on Economic Activity.

Odean, Terrance (1998), "Are Investors Reluctant to Realize their Losses", *Journal of Finance* 53, 1775-1798.

Odean, Terrance (1998), "Do Investors Trade Too Much?," *American Economic Review* 89, 1279-1298

Pagano, Marco, Fabio Panetta, and Luigi Zingales (1998), "Why do Companies Go Public? An Empirical Analysis," *Journal of Finance* 53, 27-64.

Pontiff, Jeff (1996), "Costly Arbitrage: Evidence from Closed-end funds," *Quarterly Journal of Economics* 111, 1135-52.

Quiggin, J. (1982), "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization* 3, 323-43.

Rabin, Matthew (1998), "Psychology and Economics," *Journal of Economic Literature*, 11-46.

Rabin, Matthew, and Richard Thaler (2000), "Risk Aversion," forthcoming *Journal of Economic Perspectives.*

Ritter, Jay, and Richard Warr (1999), "The Decline of Inflation and the Bull Market of 1982 to 1997," Working Paper, University of Florida.

Roll, Richard (1977), "A Critique of the Asset Pricing Theory's Tests: Part I," *Journal of Financial Economics* 4, 129-174.

Roll, Richard (1983), "Vas ist Das?," *Journal of Portfolio Management* 9, 18-28.

Roll, Richard (1986), "The Hubris Hypothesis of Corporate Takeovers," *Journal of Business* 59, 197-216

Ross, Stephen (2001), *Lectures Notes on Market Efficiency*, Sloan School of Management.

Santos, M. and Michael Woodford (1997), "Rational Asset Pricing Bubbles," *Econometrica* 65, 19-58.

Savage, L. (1964), *The Foundations of Statistics*, New York: Wiley.

Scherbina, Anna (2000), "Stock Prices and Differences of Opinion: Empirical Evidence that Stock Prices Reflect Optimism," working paper, Northwestern University.

Segal, U. (1987), "Some Remarks on Quiggin's Anticipated Utility," *Journal of Economic Behavior and Organization* 8, 145-154.

Segal, U. (1989), "Anticipated Utility: A Measure Representation Approach," *Annals of Operations Research* 19, 359-73.

Shefrin, Hersh (1999), *Beyond Fear and Greed*, Harvard Business School Press.

Shefrin, Hersh and Meir Statman (1984), "Explaining Investor Preference for Cash Dividends," *Journal of Financial Economics* 13, 253-282

Shiller, Robert (1981), "Do Stock Prices Move too Much to be Justified by Subsequent Changes in Dividends?", *American Economic Review* 71, 421-436

Shiller, Robert (1984), "Stock Prices and Social Dynamics," *Brookings Papers on Economic Activity* 2, 457-498

Shiller, Robert (2000), *Irrational Exuberance*, Princeton University Press.

Shleifer, Andrei (1986), "Do Demand Curves for Stocks Slope Down?" *Journal of Finance* 41, 579-90.

Shleifer, Andrei (2000), *Inefficient Markets: An Introduction to Behavioral Finance*, Oxford University Press.

Shleifer, Andrei, and Robert Vishny (1997), "Limits of Arbitrage", *Journal of Finance* 52, 35-55

Stein, Jeremy (1996), "Rational Capital Budgeting in an Irrational World", *Journal of Business* 69, 429-55.

Thaler, Richard (ed.), (1992), *Advances in Behavioral Finance*, Russell Sage Foundation.

Thaler, Richard (1999), "Mental Accounting Matters", Working Paper, University of Chicago.

Thaler, Richard, and Eric Johnson (1985), "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice", *Management Science* 36, 643-660.

Thaler, Richard, Amos Tversky, Daniel Kahneman, and Alan Schwartz (1997), "The Effect of Myopia and Loss Aversion on Risk-Taking: An Experimental Test", *Quarterly Journal of Economics* 112, 647-661.

Tversky, A., and D. Kahneman (1986), "Rational Choice and the Framing of Decisions," *Journal of Business* 59, 251-78.

Tversky, A., and D. Kahneman (1992), "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty* 5, 297-323.

Von Neumann, J., and O. Morgenstern (1944), *Theory of Games and Economic Behavior*, Princeton NJ: Princeton University Press.

Wurgler, Jeff, and Katya Zhuravskaya (1999), "Does Arbitrage Flatten Demand Curves for Stocks?", Working Paper, Yale University.

Yaari, M. (1987), "The Dual Theory of Choice Under Risk," *Econometrica* 55, 95-115.

Table 1: Arbitrage risks that arise in exploiting mispricing. The four risks are fundamental risk (FR), noise trader risk (NTR), implementation costs (IC) and model risk (MR).

	FR	NTR	IC	MR
Royal Dutch/Shell	$\times$	$\surd$	$\times$	$\times$
ADRs	$\times$	$\surd$	$\surd$	$\times$
Index Inclusions	$\surd$	$\surd$	$\times$	$\times$
Palm/3-Com	$\times$	$\times$	$\surd$	$\times$
Large Stock Index	$\surd$	$\surd$	$\times$	$\surd$

Table 2: Parameter values for a simple consumption-based model.

Parameter	
g_C	1.84%
σ_C	3.79%
g_D	1.5%
σ_D	12.0%
ω	0.15
γ	1.0
ρ	0.98

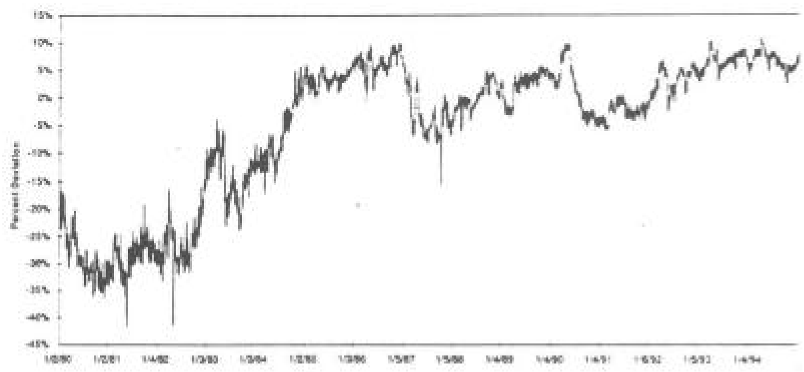

Figure 1. Log deviations from Royal Dutch/Shell parity. Source: Froot and Dabora (1999).

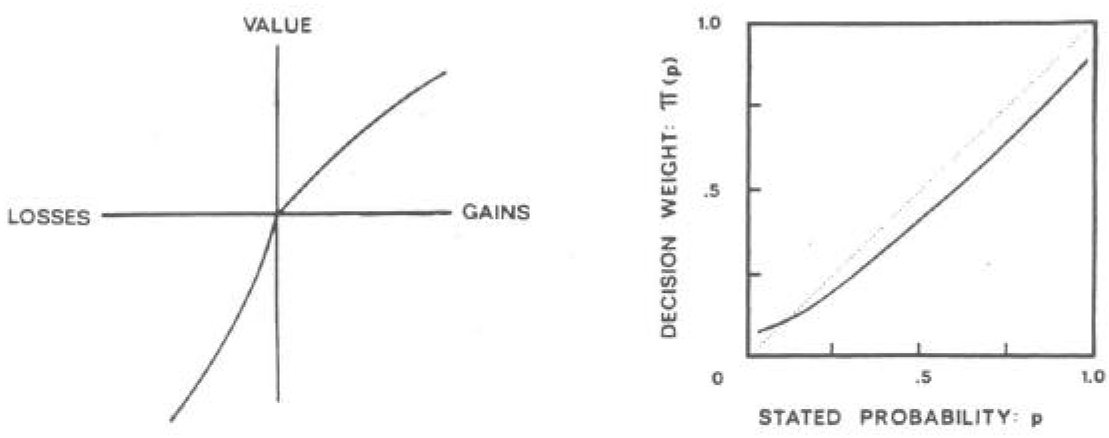

Figure 2. The two panels show Kahneman and Tversky's (1979) proposed value function v and probability weighting function π.