Three Essays on Asset Pricing, Portfolio Choice and Behavioral Finance.

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Management

by

Ehud Peleg

2008
The dissertation of Ehud Peleg is approved.

Nicholas Barberis

Richard W. Roll

Shlomo Benartzi, Committee Co-chair

Avanidhar Subrahmanyam, Committee Co-chair

University of California, Los Angeles

2008
To Gidi and Tilly.
# Table of Contents

1 Expected Returns and the Cross-Sectional Skewness of Book-to-Market Ratios .......................................................... 1  
1.1 Introduction ................................................................. 2  
1.2 Hypotheses ................................................................. 6  
1.3 Data and Methodology .................................................. 8  
  1.3.1 Construction of the Cross-Sectional Skewness Variable ..... 8  
  1.3.2 Returns and Characteristics Data ............................ 10  
  1.3.3 Sentiment Indicators .............................................. 11  
  1.3.4 Macroeconomic Conditioning Variables .................... 13  
  1.3.5 Econometric Approach ........................................... 14  
1.4 The Cross-Section of Investment Opportunities ................. 16  
  1.4.1 Future Performance of Book-to-Market Portfolios ........ 16  
  1.4.2 Predicting Size-Based Portfolios ........................... 17  
  1.4.3 \( SKEW_{bm} \) and Firm Characteristics .................... 22  
  1.4.4 Predicting the Effectiveness of Momentum and Reversals 24  
  1.4.5 Future Returns and Costs of Arbitrage .................... 26  
  1.4.6 Aggregate Predictability ..................................... 28  
1.5 Conclusion ........................................................................ 29  

2 Information, Expected Utility and Portfolio Choice .............. 52  
2.1 Introduction .................................................................... 53
3.3.4 Retirement date funds and lifecycle investment patterns . . 127

3.4 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 133

References . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 147
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>Monthly returns for portfolios based on long-term past returns, 1971-2005.</td>
<td>32</td>
</tr>
<tr>
<td>1.3</td>
<td>Monthly returns for portfolios based on short-term momentum, 1971-2005.</td>
<td>33</td>
</tr>
<tr>
<td>1.4</td>
<td>Monthly returns for portfolios based on idiosyncratic volatility, 1971-2005.</td>
<td>34</td>
</tr>
<tr>
<td>2.1</td>
<td>Consumption to wealth ratio of informed over that of the uninformed as a function of the signal spread, $\Lambda$.</td>
<td>88</td>
</tr>
<tr>
<td>2.2</td>
<td>Consumption to wealth ratio of informed over that of the uninformed as a function of risk aversion, $\gamma$.</td>
<td>89</td>
</tr>
<tr>
<td>2.3</td>
<td>Consumption to wealth ratio of informed over that of the uninformed as a function of signal noise variance, $\sigma^2_{\epsilon}$.</td>
<td>90</td>
</tr>
<tr>
<td>2.4</td>
<td>Holding in stock as a function of the propensity to consume, $\alpha$.</td>
<td>91</td>
</tr>
<tr>
<td>2.5</td>
<td>Holding in stock and in the market as a function of signal noise variance, $\sigma^2_{\epsilon}$.</td>
<td>92</td>
</tr>
<tr>
<td>2.6</td>
<td>Ex-post value of information as a function of $T$.</td>
<td>93</td>
</tr>
<tr>
<td>2.7</td>
<td>Ex-ante value of information as a function of propensity to consume, $\alpha$.</td>
<td>94</td>
</tr>
<tr>
<td>2.8</td>
<td>Initial holdings in stock, in response to private information on earnings surprise, as a function of signal noise standard deviation, $\sigma_{\epsilon}$.</td>
<td>95</td>
</tr>
</tbody>
</table>
2.9 Ex-ante value of information as a function of investment horizon, $T'$. ................................. 96

2.10 Holding in stock as a function of investment horizon, $T'$. ............... 97

3.1 Participants' Choices of the Delay between Signing-up for the Program and the Effective Date of Saving Increase ................. 134

3.2 Participants' Choices of the Month of Saving Increase .................. 135

3.3 Intended Signup Rates and the Delay between Signing-up for the Program and the Effective Date of Saving Increase ............... 136

3.4 Employee Participation Rates before and after the Implementation of Auto-Enroll Automatic Contribution Escalator .................. 137

3.5 The effect of Synchronizing Saving Increases and Pay Raises on the Degree of Interest in Joining the Program ....................... 138
# List of Tables

1.1 Summary Statistics 1970-2004. ........................................ 35  
1.2 Future Returns by SKEWbm and Book-to-Market Deciles. ..... 36  
1.3 Predicting Returns of Book-to-Market Portfolios. ............. 37  
1.4 Cross-Sectional Regressions for Size and Book-to-Market Con- 

tional on \( SKEWbm \). ............................................. 39  
1.5 Cross-Sectional Regressions for Fama-French Factor Premia Con- 

tional on \( SKEWbm \). ............................................. 41  
1.6 Time Series Regressions of SMB Portfolio, 1971-2005. ........ 42  
1.7 Time Series Regressions for Size by Book-to-Market Portfolios and  

for Size by Past-Performance Portfolios, 1971-2005. ............. 43  
1.8 Long Horizon Time Series Regressions for SMB. ............... 45  
1.9 Cross-Sectional Regressions on Firm Characteristics Conditional  
	hon \( SKEWbm \). ............................................. 47  
1.10 Time Series Regressions for Long-Short Portfolios. .......... 48  
1.11 Fama-MacBeth Regressions for Portfolios Based on Idiosyncratic  

Volatility. ....................................................... 49  
1.12 Long Horizon Time Series Regressions for CRSP Indexes, 1971- 

2005. ....................................................... 51  
2.1 Value of Insider Information About Earnings Surprises ........ 98  
3.1 Usage of Risk-Based and Retirement Date Funds ............... 139  
3.2 Equity Participation Gap ....................................... 141
3.3 Equity Market Participation .................................................. 143
3.4 Average Allocation to Equity by Age ..................................... 144
3.5 Effect of Asset Allocation Funds on Equity Allocation ............ 146
ACKNOWLEDGMENTS

I am grateful to my committee members for their excellent comments and their great support. I also thank and acknowledge the participants of the UCLA and UC Irvine finance seminars, and my colleagues, Yuzhao Zhang and Albert Sheen for great discussions on Chapter 1.

Chapter 2 is based on a co-authored working paper with Avanidhar Subrahmanyam and Jun Liu. I thank Andrew Ang, Kerry Back, Michael Brennan, Charles Cuny, Alexander David, Heber Farnsworth, Jack Hughes, Hong Liu, Venkatesh Panchapagesan, Stephen Ross, Anjan Thakor, and participants in seminars at Washington University in St. Louis, the University of California, San Diego, and at the Annual Financial Econometrics Conference at the University of Waterloo, and the conference on the Economics of Information at the University of Michigan, for valuable comments and/or discussions on that chapter.

Chapter 3 is based on a co-authored working paper with Shlomo Benzrtzi and Richard H. Thaler. I am grateful to Warren Cormier from the Boston Research Group, Jodi DiCenzo, Liat Hadar, Steve Utkus of Vanguard and Carol Waddell of T. Rowe Price for all the data and support they have provided me. I also received helpful comments from Avanidhar Subrahmanyam and Mark Grinblatt.
Vita

1970   Born, Beer Sheva, Israel


1994–1999   Chief Technology Officer and Co-Founder
             Veon, Herzeliya, Israel

1996   B.S. magna cum laude, Computer Science and Statistics
             Tel Aviv University, Israel

1999–2000   Product Manager, ECI Telecom, Petah Tikwa, Israel

2000–2001   Associate, Pitango Venture Capital, Tel Aviv, Israel

2002   Recipient, Henry Ford II Fellowship for highest GPA in class
             UCLA Anderson School
             University of California, Los Angeles, CA

2002–2008   Teaching Assistant, UCLA Anderson School
             University of California, Los Angeles, CA

2003   Recipient, J. Fred Weston Award for Academic Excellence in
             Finance, UCLA Anderson School
             University of California, Los Angeles, CA

2003   M.B.A. with Honors, UCLA Anderson School
             University of California, Los Angeles, CA

2008   Ph.D., Management, UCLA Anderson School
             University of California, Los Angeles, CA
I explore theoretical and empirical topics in portfolio choice. First, I show that the cross-sectional skewness of book-to-market ratios is an important state variable. It correlates with the business cycle and with indicators of investor sentiment. It is valuable for predicting returns on various long-short strategies, including ones based on size, volatility, book-to-market and momentum. Next, I study the consumption-investment problem of an agent with a CRRA utility, who possesses information about the future prospects of a stock. I show that information significantly alters consumption and portfolio choice. The stock holdings of informed agents are positively related to wealth, unrelated to systematic risk, and negatively related to idiosyncratic uncertainty. Furthermore, the dollar value of information depends linearly on the agent’s wealth and decreases with both the propensity to intermediate consumption and risk aversion. In the last chapter, I show that the design of retirement saving vehicles has a large effect on saving rates and investment elections, and that some of the minor details involved in the architecture of retirement plans could have dramatic effects on savings behavior.
CHAPTER 1

Expected Returns and the Cross-Sectional Skewness of Book-to-Market Ratios

The skewness of the cross-sectional distribution of book-to-market ratios is shown to be an important state variable. It varies through time in correlation with the business cycle and with indicators of investor sentiment. When times are ‘good’ a tail of low book-to-market stocks causes a left skew in the distribution. In ‘bad’ times high book-to-market stocks make the distribution more symmetric. The variable is valuable for forming conditional expectations for returns on various long-short strategies. These include ones based on firm characteristics such as size, age, volatility, profitability and dividend payment, and ones based on book-to-market and momentum. It is also instrumental in predicting the excess return on the equal-weighted market portfolio. Significant predictive ability remains even when controlling for other proxies for investor sentiment and for macroeconomic variables.
1.1 Introduction

There is a growing literature that identifies market-based proxies for times when markets are inflicted with excess investor sentiment, bullish or bearish. Lee, Shleifer and Thaler (1991) consider the discounts on close-end funds as such a proxy. Baker and Stein (2004) suggest that turnover or liquidity may proxy for sentiment in a market where individual investors face short-sale constraints. Most recently, Baker and Wurgler (2006) (BW) take an ‘operational approach’ and extract a first principal component from a number of such indicators to get a composite index.

One of the main findings of these papers is that sentiment has cross-sectional effects on mispricing and returns. That is, some stocks are more sensitive to speculative demand, and therefore have prices that diverge more from their fundamental value when ‘animal spirits’ rise. For example, when investors are optimistic, stocks that are more difficult to value or arbitrage become relatively overpriced and subsequently have lower returns.

In this chapter I show that the cross-sectional skewness of book-to-market ratios is a state variable that captures the cross-sectional effects attributed to investor sentiment. This variable’s peaks coincide with bad sentiment periods, while its troughs correspond to good sentiment periods. The idea is that when sentiment hits the market some firms experience extreme divergence of prices from fundamentals while the bulk of stock prices are more closely aligned to their ‘fair’ value. In good times we are more likely to see outlier firms with extreme high price multiples, and in bad times we are more likely to see firms at ‘fire sale’ prices. Thus the shape of the cross-sectional distribution changes with sentiment.

I show that cross-sectional skewness is correlated with some known market-
based indicators of sentiment. In particular it is correlated with the BW sentiment index and with some of its components: The number of IPOs over the last year, the dividend premium and NYSE turnover. It is also correlated with survey-based indicators of sentiment, such as the Consumer Confidence Index and the American Association of Individual Investors’ Bull minus Bear spread. Like other sentiment measures it is correlated with real economic growth variables, such as growth in industrial production and in employment.

The notion that when sentiment hits the market some firms may see a relatively extreme divergence of prices from fundamentals is closely related to the work of Shleifer and Vishny (1997) on the limits of arbitrage. They show that professional arbitrageurs, who are held accountable by their own investors, can become ineffective in extreme situations when prices diverge far from fundamentals. This is in contrast to their ability to keep other stock prices aligned with fundamentals.

Shleifer and Vishny give the 1990-1991 period as an example. Stocks of commercial banks fell sharply. Arbitrageurs trying to eliminate the mispricing invested heavily in these high book-to-market stocks. But, since the stock prices kept falling they ended up losing most of their funds under management, and were forced to liquidate their positions. Thus they could not effectively prevent the mispricing. Correspondingly, I show that the cross-sectional skewness spikes up during this period.

A good example of abnormally low cross-sectional skewness is the technology bubble. In line with the limits to arbitrage argument, Brunnermeier and Nagel (2004) document that hedge funds did not exert a correcting force on tech stocks, but rather heavily invested in them. Growth stocks had high prices, while their book values were far from stellar, leading to a ‘tail’ of very low book-to-market
ratios. This caused a negatively skewed distribution during 1999.

Previous papers use sentiment measures to predict returns of stocks that are harder to value or arbitrage. When sentiment is high these stocks are more likely to be overpriced and subsequently have low returns. When sentiment is low they are more likely to be underpriced and have a subsequent rebound. BW show that firm characteristics that proxy for difficulty in arbitrage or valuation can be used in a “conditional characteristics model.” This means that future returns depend on such characteristics. The dependence, however, varies through time according to a state variable that measures sentiment. I show that the cross-sectional skewness can be used a state variable in such a model. Moreover, it remains significant when the BW index or its components are included in the model.

One such characteristic is firm size. Small stocks are known to be prone to sentiment and to be harder to arbitrage. They are more costly to trade and sometimes impossible to sell short (D’Avolio 2002.) I find that cross-sectional skewness predicts future returns on portfolios based on size. Specifically, it predicts monthly returns on the Fama-French SMB portfolio even when controlling for other known predictors, such as the risk-free, the term-spread and the default spread. The effect is not driven by the January effect and is robust to the inclusion of macroeconomic variables and investor sentiment proxies.

Returns of portfolios formed on other characteristics such as age, volatility, dividend payment and profitability are also predictable, even when controlling for the three Fama-French factors and the momentum factor. The effect also persists when controlling for firm size and book-to-market directly rather than the linear factors. The direction of predictability is that high skewness leads to high returns for firms that are young, volatile, unprofitable or that do not pay
Following Wurgler and Zhuravskaya (2002) I consider the idiosyncratic volatility of stocks as a proxy for the costs of arbitrage. I show that following low skewness (corresponding to good sentiment periods) firms with higher idiosyncratic volatility underperform those with low volatility, but following high skewness (corresponding to bad sentiment periods) they significantly outperform them. This is in line with the limits to arbitrage theory that such stocks become mispriced when sentiment reigns in the market.

Expected returns on portfolios based on past performance, i.e. long-term reversal and short-term momentum, are also shown to correlate with the lagged skewness variable. Long-term reversal strategies are most effective following high skewness periods (‘bad sentiment’) when losers show subsequent good returns. However, they are not effective following periods of low skewness (‘good sentiment’) when long-term losers continue to underperform. Following low skewness periods, returns to short-term momentum strategies result almost solely from shorting losers, as last year’s winners do not outperform the market. Coming out of a ‘bad’ period short-term winners show strong momentum and the ‘long’ leg of the strategy is the profitable one.

I investigate whether cross-sectional skewness predicts aggregate market returns. I find that like other sentiment indexes it predicts future excess returns on the equal-weighted CRSP index. A one standard deviation increase in skewness predicts an increase of about five percentage points in the subsequent annual return of the index. However, the skewness is negatively correlated with future returns on the value-weighted index, and it does not predict it significantly.

The next section presents the main hypotheses of the chapter. Section 1.3 introduces the data and the skewness measure. Section 1.4 looks at the cross-
sectional and aggregate predictability using this variable. Section 1.5 concludes.

1.2 Hypotheses

Hypothesis I - Cross-sectional skewness of book-to-market ratios is correlated with ‘sentiment’ measures.

The first hypothesis is that the shape of the cross-sectional distribution of book-to-market ratios changes through time. In periods when investors are bullish there is a left tail in the distribution and the skewness drops. When investors are bearish some stock prices get excessively depressed and the skewness rises.

Suppose investors are overly bullish about the market. Why would we expect their excess demand for stocks to alter the shape of the cross-sectional distribution of book-to-market ratios? I offer three explanations for why a group of stocks will get disproportionately overpriced and lower the skewness of the distribution. Mirror arguments can explain why a bear market will be related to high skewness.

First, investors tend to focus on certain types of stocks at different points in time for rational or irrational motives. Barberis and Shleifer (2003) propose that investors classify risky assets into ‘styles,’ and move funds back and forth into them as a group. This can lead to circumstances where much of the excessive demand is directed at one style, causing it to be relatively mispriced. For example, “New Economy” stocks in the late 90s undoubtedly contributed to the cross-sectional skewness at that point.

Second, suppose investors are exposed to good news (signals) about stocks and they are overconfident about the precision of these signals. If they face fixed costs for trading on these signals, they only trade in the stocks for which they assess their potential profits to be higher than the costs. In that case only stocks
for which signals are higher than a certain threshold become mispriced.

The third explanation is related to the limits of arbitrage. It could be that the rational arbitrageurs are the ones facing costs or limits to their arbitrage ability. Shleifer and Vishny (1997) consider professional arbitrageurs who manage other people’s capital. Such arbitrageurs identify opportunities, take contrarian positions in the mispriced stocks and wait for prices to move back to fundamentals. If prices of the assets in which they are invested diverge much further from fundamentals the arbitrageurs lose money, their investors withdraw funds and they are forced to liquidate their positions. Thus, in equilibrium, arbitrageurs prefer to avoid opportunities for which they expect mispricing to significantly worsen and they become most ineffective in extreme situations when prices diverge far from fundamentals. In up-markets we can expect ‘extreme’ overpricing situations and in down-markets ‘extreme’ underpricing situations.

Hypothesis II - Cross-sectional skewness is positively correlated with future returns on stocks that are more difficult to value or arbitrage.

As previously mentioned, sentiment indexes are instrumental in predicting conditional stock returns based on firm characteristics. Lee et al. (1991) consider small stocks because they are disproportionately held by individuals. BW consider a wider classification. They hypothesize that sentiment will affect stocks that are more costly to arbitrage or more difficult to value, and suggest young, small, volatile and unprofitable firms as such stocks.

A testable hypothesis is that skewness is positively correlated with future returns of small, young and volatile stocks. For example low skewness implies overvaluation and subsequent low returns for small stocks. In testing this hypothesis it is crucial to show that the effect is not driven solely by firms that have extreme book-to-market ratios, and to show that skewness predicts returns
on small firms regardless of their book-to-market ratios.

Skewness is also hypothesized to predict the source of positive returns to long-short strategies like value and momentum. Following high skewness, high book-to-market stocks are expected to rebound, and so are long-term losers. Following low skewness the value strategy shows good returns because of the short position in ‘growth’ stocks that under-perform the market, while high book-to-market firms will not outperform the median book-to-market firms.

*Hypothesis III - Cross-sectional skewness is positively correlated with future excess returns on the equal-weighted market index.*

Since skewness positively predicts the returns on small firms it is expected to positively predict the returns on the equal-weighted CRSP index and negatively predict, if at all, returns on the value-weighted index. This is because the mispriced stocks are expected to be under-weighted in the value-weighted index.

### 1.3 Data and Methodology

#### 1.3.1 Construction of the Cross-Sectional Skewness Variable

The main variable used in this chapter is the cross-sectional skewness of log book-to-market ratios \( SKEW_{bm} \). I use data from COMPUSTAT and CRSP to construct it. Each period skewness is measured as:

\[
SKEW_{bm_t} = \frac{N}{(N - 1)(N - 2)} \sum_{i}^{N} \left( \frac{bm_{i,t} - \overline{bm_t}}{\sqrt{S_t^2}} \right)^3
\]

where \( bm_{i,t} \) is the log of the book-to-market ratio of firm \( i \) at time \( t \). \( \overline{bm_t} = \frac{\sum bm_i}{N} \) is the sample cross-sectional mean of \( bm_{i,t} \) at time \( t \) and \( S_t^2 = \frac{1}{N-1} \sum (bm_i - \overline{bm})^2 \) is the sample variance. I include only ordinary common shares of firms incorporated in the US (share code 10 or 11) that are listed on NYSE, AMEX or NASDAQ.
Book-to-market ratios are calculated as follows. Book values are taken from the COMPUSTAT quarterly series. Book data are dated by the calendar quarter in which the fiscal quarter ends. They are matched to market values skipping one full calendar quarter to make sure that the book data are already publicly available. For example, a book value for a fiscal quarter ending in March is matched to market values at the end of July, August and September.

In an unreported analysis I use the time-series that is based on annual COMPUSTAT data, where firms’ book-to-market ratios are calculated using the convention of Fama and French (1992). That is, I match book values for fiscal year-ends in calendar year \( t - 1 \) to market values from July \( t \) to June \( t + 1 \). The series includes additional years of data, but has large discontinuities every twelve months. Most of the results remain similar to those reported here with the quarterly series.

Figure 1.1 plots the time-series of \( SKEW_{bm} \). It is negative for the most part and it is within a range of values from -1.32 to 0.12. The mean is -0.72 and the standard deviation is 0.27. Some low points coincide with known periods of high sentiment like the early 80s and the hi-tech bubble. The late 60s show up as high sentiment in the unreported series that uses annual book values. Some of the variable’s highs coincide with low sentiment: the mid 70s, late 80s, the 1990-1991 recession, and the burst of the bubble after 2000.

Previous papers, such as Kothari and Shanken (1997), document that aggregate book-to-market ratios, like the DJIA aggregate book-to-market ratio, have pronounced trends and are close to unit root. Skewness has a lower one-lag serial correlation, 0.90. The augmented Dickey-Fuller statistic is -3.96, rejecting non-stationarity at p-value 0.01.
1.3.2 Returns and Characteristics Data

The empirical tests center on the relationship between stock returns and firm characteristics. The accounting data for these are taken from the annual COMPUSTAT database. I follow Fama and French (1992) and match accounting data for fiscal year-ends in calendar year $t - 1$ to monthly returns from July $t$ to June $t + 1$.

Size, $ME$, is the market equity of the firm computed as price times shares outstanding from CRSP. Firm size from June of year $t$ is matched to returns from July of year $t$ to June of year $t + 1$. Age is the number of years since the firm’s first appearance on CRSP. Volatility, $\sigma$, is the standard deviation of monthly returns over the twelve months ending in December of year $t - 1$, and is matched to returns from July of year $t$ to June of year $t + 1$.

Earnings, $E$, are defined as income before extraordinary items plus income statement deferred taxes minus preferred dividends. The profitability dummy, $E > 0$, takes a value of one if earnings are greater than zero. $ROE$ is zero for unprofitable firms and is equal to $E/BE$ for profitable firms, where $BE$, is shareholders’ equity plus deferred taxes.

Dividend to equity, $D/BE$, is determined from COMPUSTAT as dividend per share at ex date times share outstanding divided by book equity. A firm is a dividend payer, $D > 0$, if COMPUSTAT reports positive dividend per share by the ex date.

Two proxies for asset tangibility are considered following BW: property, plant and equipment over assets, $PPE/A$, and research and development expense over assets, $RD/A$, when it is positive.
1.3.3 Sentiment Indicators

In order to show the incremental value of $SKEWbm$, I wish to control for other known proxies for sentiment. Previous literature suggests many possible alternatives. BW review the literature and suggest a composite index. I use it as the main control variable for sentiment. It is a combination of six sentiment indicators, which I also consider separately. These are: the closed end fund discount, number and average first-day returns on IPOs, the dividend premium, the equity share in new issues, and NYSE share turnover. End-of-year data for the index and the components are taken from Jeffrey Wurgler’s Web site. The definitions of the variables are given in BW. I briefly review their definitions below. The correlations with $SKEWbm$ are given in Table 1.1.

The closed-end fund discount, CEFD, is the average difference between the net asset values of closed-end stock fund shares and their market prices. Prior work suggests it is inversely related to sentiment. It is not significantly correlated to $SKEWbm$. However, papers like Qiu and Welch (2004) already show that it has lost its correlation to investor sentiment in later periods.

NYSE share turnover is based on the ratio of reported share volume to average shares listed on NYSE. The variable, TURN, is defined as the natural log of the raw turnover ratio, detrended by the five-year moving average. According to Baker and Stein (2004) turnover or liquidity may proxy for sentiment in a market where individual investors face short-sale constraints. It is negatively correlated to $SKEWbm$ as expected.

The number of IPOs in the last year, NIPO, is correlated to $SKEWbm$. This is expected if more firms IPO when multiples for comparable firms are relatively high. Mechanically, if IPOs tend to have low book-to-market ratios then their appearance in the cross-section can cause a left skew. To control for such an effect
I also construct a measure of skewness that includes only firms with more than one year of observations in CRSP. The results throughout the chapter still hold. The average first-day returns on last year’s IPOs, RIPO, has a lower correlation with $SKEW_{bm}$.

The equity share in new issues, $S$, is defined as gross equity issuance divided by gross equity plus gross long-term debt issuance using data from the *Federal Reserve Bulletin*. I do not find it to be correlated with $SKEW_{bm}$. The last of these proxies is the dividend premium, PDND, which is the log difference of the average market-to-book ratios of payers and nonpayers. It is significantly correlated with $SKEW_{bm}$.

In sum, $SKEW_{bm}$ is correlated with a number of market-based sentiment indicators. It is most highly correlated with the dividend premium and the number of IPOs. It is also significantly correlated with the BW sentiment index. Throughout the chapter I would like to determine the effects of $SKEW_{bm}$ that are not captured by this index. To accomplish this, I use the residuals from a monthly contemporaneous regression of $SKEW_{bm}$ on the index as an explanatory variable.

I also consider the correlation of $SKEW_{bm}$ to measures formed directly from investor surveys. Qiu and Welch (2004) consider the Michigan Consumer Confidence Index as a measure of investor sentiment. This measure is available in monthly intervals since 1978. I find that $SKEW_{bm}$ has a significant -0.14 correlation with that index, while the series differences have a -0.17 correlation.

The American Association of Individual Investors’ (AAII) sentiment survey measures the percentage of individual investors who are bullish, bearish, and neutral on the stock market short term. These individuals are polled from the AAII Web site on a weekly basis. $SKEW_{bm}$ has a significant 0.18 correlation
with the monthly average of the bull minus bear spread measure of that survey.

1.3.4 Macroeconomic Conditioning Variables

Chen, Roll and Ross (1989) and Ferson and Harvey (1999) document that macroeconomic variables that predict aggregate stock and bond returns through time, also provide significant cross-sectional explanatory power for stock portfolio returns. I therefore look at the relationship between $SKEW_{bm}$ and some of these variables.

The first is the dividend price ratio, DPR, which is the difference between log of the S&P 500 and the sum of dividends paid over the previous twelve months. Since sentiment is expected to be related to the equal-weighted portfolio, I also look at EBM, the equal-weighted average of book-to-market ratios of CRSP firms.

Next I consider three bond-related variables, which are known to capture the macroeconomic state. The first is the three-month treasury bill, TBL. It is taken from the 3-month Treasury Bill: Secondary Market Rate series of the economic research data base at the Federal Reserve Bank at St. Louis (FRED). The two others are: the default yield spread, DFY, which is the difference between BAA- and AAA-rated corporate bond yields, and the term spread, TMS, which is the difference between the long-term yield on government bonds and the T-bill.

I also control for three direct measures of the real activity in the economy: an indicator of recession as defined by NBER, growth in employment which is defined as the log growth in Total Seasonally-Adjusted Nonfarm Employment from the BLS Web site, and growth in industrial production from the Federal Reserve Board G.17 release.

As can be seen in Table 1.1, $SKEW_{bm}$ is negatively correlated with growth in employment and growth in industrial production. Also, some of its identifiable
peaks in Figure 1.1, like 1991 and 2001, correspond to NBER recessions.

### 1.3.5 Econometric Approach

The chapter aims at relating $SKEW_{bm}$ to overvaluation or undervaluation of certain stocks. In particular, it focuses on predicting the returns due to the correction of such mispricing. To allow for a reversal pattern I look at the relationship between monthly stock returns and one-year lagged $SKEW_{bm}$. As an alternative I also consider predictive regressions for long-term portfolio returns with lengths of twelve to sixty months.

Two regression approaches are used. One is a conditional characteristics model as in BW, where monthly returns on portfolios based on firm characteristics are predicted by one-year lagged $SKEW_{bm}$. A typical regression model is of the following form

$$R_{H,t} - R_{L,t} = \beta_0 + \beta_{SK}SKEW_{bm_{t-year}} + \gamma \cdot Controls_t + \epsilon_t \quad (1.1)$$

$R_{H,t}$ and $R_{L,t}$ are respectively monthly returns on equal-weighted portfolios of the top three deciles and the bottom three deciles, which are constructed for different characteristics based on NYSE breakpoints. The characteristics I consider are firm size, age, volatility, ROE, indicators for profitability and dividend payment, and asset tangibility ratios: PPE/A and R&D/A. The period in consideration is 1971-2005. In the case of size, Fama-French SMB is used instead of the High minus Low portfolio.

The contemporaneous controls are returns on the factors of a four-factor model. They include RMRF, SMB, HML and UMD. RMRF is the excess return on the value-weighted market portfolio. SMB is the return on a portfolio long small firms and short big firms controlling for book-to-market. HML is the
return on a portfolio of high book-to-market firms minus that of low book-to-market firms controlling for size. UMD is the return on high momentum minus low momentum stocks, where momentum is defined over months [-12,-2]. Returns are taken from Ken French’s Web site and the construction of the portfolios is described there. SMB is not included as an explanatory variable when it is the dependant variable.

The second econometric approach follows Ferson and Harvey (1991). It tests whether cross-sectional premia for certain characteristics or factor loadings are predictable. In the first stage I run a cross-sectional regression each month of the form:

\[ R_{i,t}^e = \gamma_{0,t} + \gamma_t \cdot X_{i,t-1} + \epsilon_{i,t} \]  

(1.2)

\( R_{i,t}^e \) is the excess return on a stock or a portfolio. \( X_{i,t-1} \) is either a firm characteristic known prior to time \( t \), or a factor loading calculated using information prior to time \( t \). \( \gamma_t \) is a vector of premia in the cross-section for the components of \( X_{i,t-1} \). In the second stage I regress components of \( \gamma_t \) on \( SKEWbm_{t-year} \) to see whether the cross-sectional relationship between returns and the components of \( X_{i,t-1} \) is predictable using the skewness measure. This regression is of the form:

\[ \gamma_{i,t} = \beta_0 + \beta_{SK} \cdot SKEWbm_{t-year} + \epsilon_t \]  

(1.3)

I run these regressions for several stock characteristics controlling for size and book-to-market in the cross-section.
1.4 The Cross-Section of Investment Opportunities

1.4.1 Future Performance of Book-to-Market Portfolios

I first test whether at least some of the reversion of $SKEW_{bm}$ from its peaks and troughs comes about through market returns rather than book returns. Namely, when there is skewness to the left, low book-to-market firms are due to have subsequent low returns. When skewness is to the right, high book-to-market firms are expected to have high returns. This will support the notion that the changes in skewness are at least partly related to temporary shocks to stock prices.

I break down the period of 1970-2004 to low skewness periods defined as the lowest quartile, mid-skewness periods, which fall in the interquartile range, and the highest quartile of skewness periods. Table 1.2 shows that following low instances of $SKEW_{bm}$ most of the value premium is due to the lower performance of the low book-to-market deciles versus the mid-range deciles. Following high $SKEW_{bm}$ it is due to the better performance of the high book-to-market portfolios. The different states, however, do not show a difference in the total value spread.

This suggests that the value strategy of buying high book-to-market and selling low book-to-market, can be broken down to high minus mid, and mid minus low. The performance of the former should be positively related to skewness, while that of the latter should be negatively related to it. To test this, I form monthly portfolios by NYSE deciles of book-to-market from 1971-2005. Two strategies are constructed: an equal-weighted portfolio of the top three deciles minus that of the mid four deciles, and an equal-weighted portfolio of the mid four deciles minus that of the lowest three deciles.
Table 1.3 shows the results of running the regressions for predicting the portfolios with one-year lagged $SKEWbm$. The regression equations are:

$$R_{H,t} - R_{M,t} = \beta_0 + \beta_{SK} SKEWbm_{t-\text{year}} + \gamma \cdot Controls_t + \epsilon_t$$  \hspace{1cm} (1.4)

$$R_{M,t} - R_{L,t} = \beta_0 + \beta_{SK} SKEWbm_{t-\text{year}} + \gamma \cdot Controls_t + \epsilon_t$$  \hspace{1cm} (1.5)

The regression is run once without controls. Then, I control for the three Fama-French factors and the momentum factor. The results suggest that such a reversal relationship does exist, and that $SKEWbm$ predicts each portion of the value spread even when controlling for concurrent HML. The coefficient is still significant for the High-minus-Neutral portfolio when $SKEWbm$ is orthogonalized with respect to the BW sentiment index. But, it loses significance for the Neutral-minus-Low portfolio in that case.

1.4.2 Predicting Size-Based Portfolios

Sentiment is hypothesized to have an effect on the relationship between returns and firm size. Small firms have been mentioned in the past as harder to arbitrage and as more prone to investor sentiment (e.g. Lee et al., 1991.) The goal of this section is to show that the skewness of the cross-sectional distribution of book-to-market ratios is instrumental in forming expectations about the future performance of firm-size based strategies.

I consider the cross-section of equal-weighted returns on twenty five portfolios based on ME and BE/ME from Ken French’s Web site. I use the portfolios in two Fama-MacBeth settings. One in which I estimate the slopes on the size and book-to-market characteristics and the other where I measure the risk premium or the slope on betas with each of the Fama French factors. An unreported analysis using the value-weighted returns of the portfolios yields similar results.
First, each month I run the cross-sectional regression in equation (1.6).

\[ R_{i,t} - R_{f,t} = \beta_{0,t} + \beta_{ME,t} \cdot \log(ME)_{i,t-1} + \beta_{BM,t} \cdot \log(B/M)_{i,t-1} + \epsilon_{i,t} \] (1.6)

Table 1.4 reports the mean and the Fama-MacBeth t-stat for the full sample and for the subsamples as described in the previous section. The size coefficient is only negatively significant following the high \( SKEW_{bm} \) period. It is positive but insignificant following the low period. The difference in means between these two periods is 0.43 and is significant at 0.01.

I run a regression for the coefficient on size versus previous year’s \( SKEW_{bm} \).

\[ \beta_{ME,\tau} = \gamma_0 + \gamma_{SK} \cdot SKEW_{bm_{\tau-year}} + \epsilon_{\tau} \] (1.7)

Results are given in Panel B of Table 1.4. T-stats are based on White (1980) (Newey-West errors have similar p-values.) \( SKEW_{bm} \) is significant in predicting the cross-sectional coefficient for size.

As an alternative to the characteristic-based specification, I perform Fama-MacBeth type regressions where in each cross-section portfolio returns are regressed on their betas with each of the Fama-French factors. Betas are calculated using monthly data for the previous five years. The coefficients in these regressions are estimates of the cross-sectional risk-premia for each of the factors. The results are reported in Table 1.5.

The ANOVA t-test for difference in means for \( \lambda_{SMB} \) between periods following high and low \( SKEW_{bm} \) is significant at 0.01. Positive loading on SMB earns a significant positive premium only after high \( SKEW_{bm} \). During the period when the premium on SMB is significant, the premia on the two other factors are not. The premium on \( RMRF \) increases with \( SKEW_{bm} \), which is an indication that high skewness (correlated with ‘bad’ sentiment) is associated with higher expected market premium. Although, like other studies, this premium is found to
be negative in a three-factor model. The premium on HML disappears following high skewness indicating that size might be more effective in explaining the cross-section of expected returns coming out of a ‘bad’ period.

To quantify the effect on returns and control for other predictors of SMB, I run a regression of the monthly portfolio returns on one-year lagged \( SKEW_{bm} \). The results are in Table 1.6. In a univariate regression, \( SKEW_{bm} \) is significant in predicting SMB at 0.01 level. The economic magnitude of the effect is that one standard deviation increase in \( SKEW_{bm} \) predicts an additional 0.5 percentage point per month return on the SMB portfolio.

The cross-sectional skewness remains significant even when the one-year lagged log of the equal-weighted mean of book-to-market ratios is included. It is also significant when additional bond related variables are added. These variables are lagged only one quarter. They are the t-bill, the term spread and the default spread. The skewness is still significant when additional macroeconomic variables are added. These are growth in industrial production, growth in employment and an NBER recession dummy. Furthermore, \( SKEW_{bm} \) remains significant when orthogonalized with respect to the BW sentiment index.

1.4.2.1 Results for Size Controlling for Book-To-Market and Past Performance

In order to verify that SMB predictability is not due to extreme growth or value firms, I perform the time series regression controlling for book-to-market quintile. Panel A of Table 1.7 considers five regressions. One-year lagged \( SKEW_{bm} \) is the regressor in all of them. The dependent variable is one month return for a portfolio of the smallest quintile of firms in a given book-to-market quintile minus the return on the largest quintile of firms in the same book-to-market quintile.
Portfolios and returns are from Ken French Web site.

The regressions show that results are not driven just by stocks of the extreme quintiles in book-to-market and predictability is across-the-board. Although the coefficients are slightly higher for the lowest and highest book-to-market quintiles. As in previous cases, the significance and magnitude of the $SKEW_{bm}$ coefficient is not materially altered when the monthly BW sentiment index is included in the regression.

A different concern is that the predictive results are related to the long-term reversal effect documented by De Bondt and Thaler (1985). They show that contrarian strategies which are long past long-term losers and are short past long-term winners earn abnormal returns. Panel B of Table 1.7 considers five regressions where the dependent variable is one month return for a portfolio of the smallest quintile of firms in a given long-term reversal quintile minus the return on the largest quintile of firms in the same long-term reversal quintile. The long-term portfolios are formed monthly on prior (13-60) returns. Return breakpoints are based on NYSE quintiles. Portfolios and returns are again taken from Ken French’s Web site.

The regressions show that the predictability of small-minus-big by $SKEW_{bm}$ is not merely a long-term reversal effect. $SKEW_{bm}$ predicts the returns of small-minus-big in each long-term quintile. The coefficient is however largest and most significant when past long-term losers are considered.

Although there is no clear reason why the results would be driven by short-term momentum, the last panel in Table 1.7 shows that the predictability of size portfolios is not driven by short-term momentum. The method is similar to that used to control for long-term reversal.
1.4.2.2 Long-Run Analysis

A one-year lag may not be the correct time-frame to capture the reversal effect. I perform time-series prediction analysis for the SMB portfolio of varying lengths. Log returns of SMB over twelve to sixty months are regressed on the one-month lagged skewness variable.

\[ r_{t,t+K} = \beta_0 + \beta_{SK} \cdot SKEWbm_{t-1} + \epsilon_t \]  \hspace{1cm} (1.8)

Table 1.8 reports the results. T-stats are based on Newey-West standard errors with q=K-1. Portfolio returns are predictable at 0.01 for all time periods. The R-squares increase with the length of the period up to three years and then remain flat. The economic significance of the results is that a change of one standard deviation of \( SKEWbm \) upwards will increase the annual return on the SMB portfolio by about six percentage points. This drops to about four percentage points per year after three years.

In Panel B of the table the one-year regression is reported with controls for other predictors. \( SKEWbm \) remains highly significant. The controls include the equal-weighted mean of log book-to-market ratios of the same universe of firms as that included in \( SKEWbm \). Also included in the regression are the first principal measure of BW and its components.

Although the BW index predicts SMB by itself. This predicting ability disappears when \( SKEWbm \) is included. In the presence of \( SKEWbm \) the number of IPOs and the dividend premium remain significant in a regression that includes all six components. These controlled regressions indicate that \( SKEWbm \), although correlated with known market-based indicators, still provides new information for predicting small stock returns. Similar results hold when I use the BW method of using end-of-year values of the market-based indicators to predict
monthly returns throughout the year.

1.4.3 \textit{SKEWbm} and Firm Characteristics

Different stock characteristics may help identify stocks that are more susceptible to sentiment or that have higher costs of arbitrage. These characteristics will identify the stocks that will be overpriced during the periods of high sentiment, and underpriced when there is low sentiment. Therefore, if we condition on sentiment, we can use them to predict subsequent reversal patterns. To test whether this is the case I look at the predictability of their cross-sectional ‘premia.’ I first run monthly cross-sectional regressions of the form:

\begin{equation}
R_{e,i,t} = \beta_{0,t} + \beta_{ch,t} \cdot \text{Char}_{i,t-1} + \beta_{BM,t} \cdot \log(B/M)_{i,t-1} + \epsilon_{i,t} \tag{1.9}
\end{equation}

and

\begin{equation}
R_{e,i,t} = \beta_{0,t} + \beta_{ch,t} \cdot \text{Char}_{i,t-1} + \beta_{BM,t} \cdot \log(B/M)_{i,t-1} + \beta_{ME,t} \cdot \log(ME)_{i,t-1} + \epsilon_{i,t} \tag{1.10}
\end{equation}

\text{Char} is one of the following: an indicator whether the firm paid dividends, an indicator whether it was profitable in the previous fiscal year, the firm’s age (years in CRSP) or the ratio of R&D to Assets in the last fiscal year.

I then run a second stage regression:

\begin{equation}
\beta_{ch,\tau} = \gamma_0 + \gamma_{SK} \cdot \text{SKEWbm}_{\tau-year} + \epsilon_{\tau} \tag{1.11}
\end{equation}

Table 1.9 provides the estimated coefficients $\gamma_0$ and $\gamma_{SK}$, for the period 1971-2005. Since \textit{SKEWbm} has been standardized to have zero mean the intercept represent the full sample means of the $\beta$s from the first stage. Dividend payers and profitable firms do not significantly underperform or outperform the other firms in the full time-series sample. But their performance relative to non-payers and
non-profitable firms improves as the skewness decreases. For example, one year after high sentiment period (characterized by $\text{SKEWbm}$ equal to mean minus one standard deviation) the future performance of payers is expected to be 0.54 percentage points per month higher than non-payers. Following low sentiment (characterized by $\text{SKEWbm}$ equal to mean plus one standard deviation) payers are expected to underperform non-payers by 0.86 percentage points per month. Some of the significance of the effect as well as its magnitude decrease when size is included in the cross-sectional regressions.

I further explore the role of conditional pricing by looking at predicted returns of portfolios based on equal-weighted portfolios of High minus Low values of various firm characteristics as in equation (1.1). The ‘High’ portfolio includes firms in the top three deciles based on NYSE breakpoints, and the ‘Low’ portfolio includes firms in the bottom three deciles based on NYSE breakpoints.

Table 1.10 considers the predictability for portfolios based on a number of characteristics that have previously been identified as related to difficulty in valuation or arbitrage. When skewness is high, young, volatile, non-earners and non-payers are expected to outperform. This is significant even when controlling for the three Fama-French factors and the Momentum Factor. They are also significant when $\text{SKEWbm}$ is orthogonalized with respect to the BW index.

Returns on portfolios formed on proxies for asset tangibility, i.e. PPE/A, and R&D/A are also predictable. But, the predictability goes away when controlling for the contemporaneous factors. High skewness, i.e. bad sentiment, predicts good returns on low PPE/A and high R&D/A firms.
### 1.4.4 Predicting the Effectiveness of Momentum and Reversals

In this section I investigate the relationship between $SKEW_{bm}$ and the subsequent return profiles for two strategies based on relative past performance: long-term reversal (De Bondt and Thaler 1985) and short-term momentum (Jegadeesh and Titman 1993.) I wish to expand on previous papers that show that the returns to portfolios formed on past performance vary with the business cycle. Ball, Kothari and Shanken (1995) point out that positive returns on long-term reversals are mostly due to good performance of losers and occur predominantly after down markets. Chordia and Shivakumar (2002) show that returns to short-term momentum vary through the business cycle and can be explained by a set of lagged macroeconomic variables.

I first turn to long-term reversals and show that positive returns to the strategy indeed follow high $SKEW_{bm}$ periods. This is shown in Figure 1.2. It shows the returns to each of the long-term reversal deciles following three $SKEW_{bm}$ periods: low $SKEW_{bm}$ (‘good period’), mid $SKEW_{bm}$, and high $SKEW_{bm}$ (‘bad period’). The ten equal-weighted decile portfolios are based on long-term past returns (13-60) where breakpoints are set by NYSE firms.

The figure supports the Ball et al. (1995) findings. Following periods of high skewness losers outperform winners. In fact, there is a monotonous relationship along the deciles. On the other hand following periods of low skewness losers do not perform as well as winners. Namely, firms that underperformed during periods of good sentiment go on to underperform rather than to experience reversal.

To verify the significance of the effect, I consider a regression of the long-term reversal factor, which measures the returns to a long-term reversal strategy, on $SKEW_{bm}$. The factor returns are taken from Ken French’s Web site. It is based on six value-weighted portfolios formed on size and prior (13-60) returns. The
portfolios, which are formed monthly, are the intersections of 2 portfolios formed on size and 3 portfolios formed on prior (13-60) return. Breakpoints are based on NYSE percentiles. The factor is the average return on the two low prior return portfolios minus the average return on the two high prior return portfolios.

I find that returns to the reversal strategy significantly increase with $SKEW_{bm}$ even when controlling for the January effect and for the four factors. One standard deviation increase in $SKEW_{bm}$ leads to an additional 0.33 percentage points per month to the reversal strategy, which overall averages 0.31 percentage points per month over the 1971-2005 period. If reversal is indeed related to overreaction, as De Bondt and Thaler (1985) suggest, then it is overreaction to bad news rather than good news. The presence of such overreaction can be detected by the form of the cross-sectional distribution of book-to-market ratios.

I now turn to the short-term momentum strategy. Although the short-term momentum strategy seems to be profitable on average in all periods. Figure 1.3 provides evidence about the source of the profits for this long-short strategy. Every month ten equal-weighted portfolios are formed based on stocks short-term past returns (2-12). As in Figure 1.2, the monthly returns are grouped into three periods according to the lagged $SKEW_{bm}$ variable. Portfolios and monthly returns are taken from Ken French’s website.

While on average both sides of the long-short strategy contribute to the returns, it appears that coming out of a bad period (i.e. high $SKEW_{bm}$) continuing relative out-performance of winning portfolios are the source of the positive returns. On the other hand, subsequent to low $SKEW_{bm}$ periods momentum profits are mostly due to the continuing underperformance of losers relative to the average.

To verify the significance of these results I use a regression framework where
I predict the excess monthly returns of the short-term top decile (winners) over the market, and the excess returns of the market over the short-term bottom decile (losers). $SKEWbm$ predicts both of those spreads significantly. It remains so when the four-factors including momentum are included and when it is orthogonalized to the BW index.

The magnitudes implied by the regressions are not small. One standard deviation increase in $SKEWbm$ leads to a 0.73 percentage points increase in monthly excess return of the winner portfolio over the market. One standard deviation decrease in $SKEWbm$ leads to 1.05 percentage points increase in the spread of the market over the loser portfolio.

The findings support a notion that momentum should not be seen as a monolithic strategy. In a recuperating economy its source is in certain stocks that see runs of positive returns. This is in line with the limits to arbitrage argument that some stocks were more mispriced during the down market, but also requires that the correction will not be instantaneous but take about a year. At other times, like during a collapsing bubble, positive returns are a result of certain stocks seeing runs of negative returns.

1.4.5 Future Returns and Costs of Arbitrage

Wurgler and Zhuravskaya (2002) suggest that a stock’s costs of arbitrage can be measured as the historical variance of a zero-net-investment portfolio that holds one dollar long in the stock and one dollar short in a portfolio of substitutes. They consider either the market portfolio or three stocks matched by industry, size and book-to-market as substitutes and show that for the 259 stocks in their sample the methods yield highly correlated results. Since I have to estimate costs for the entire cross-section and over a longer period, I opt for the market approach
and use CAPM idiosyncratic volatility as a proxy for arbitrage risk.

Idiosyncratic volatility (ivol) is calculated as the standard deviation of the residuals in the regression of the excess stock return versus the excess market return. Every month, I use the previous five years to estimate each stock’s idiosyncratic volatility as the standard deviation of $\epsilon_{i,t}$ in equation (1.12).

$$R_{i,t} - R_{f,t} = \beta_{0,i} + \beta_i(R_{m,t} - R_{f,t}) + \epsilon_{i,t} \quad (1.12)$$

Next, I form ten monthly portfolios by ranking the stocks according to their idiosyncratic volatility. Figure 1.4 shows the equal-weighted average returns in excess of the market of these portfolios. The three lines represent months following low $SKEW_{bm}$ defined by the lowest 25-percentile of the 1970-2004 time-series, following mid $SKEW_{bm}$ which is the interquartile range and following the highest 25-percentile. The figure indeed shows that following positive skewness high volatility stocks will perform better than low volatility stocks, and vice versa following negative skewness. A limits to arbitrage argument would be that during the high sentiment period which drove the low skewness, high ivol stocks became overpriced because arbitrageurs could not undo the mispricing.

Since the hypothesis is about the time-series of the cross-sectional slope of returns versus idiosyncratic volatility I use the Fama-MacBeth framework to test it. Each month in the sample of 1971-2005, I run a cross-sectional regression of the excess stock return on the idiosyncratic volatility. Stocks with less than five years of data are excluded. Every month the regression in equation (1.13) is estimated.

$$R_{i,t} = \beta_{0,t} + \beta_{ivol,t}t \cdot ivol_{i,t-1} + \beta_{ME,t} \cdot \text{log}(ME)_{i,t-1} + \beta_{BM,t} \cdot \text{log}(B/M)_{i,t-1} + \epsilon_{i,t} \quad (1.13)$$

To show the correlation to lagged $SKEW_{bm}$, the sample is first broken down to three periods according to lagged $SKEW_{bm}$ and a test for the difference
in means is conducted. Table 1.11 considers the sample and subsamples and provides the mean and Fama MacBeth t-stats for the intercept and the slope of the volatility. The full sample coefficient on ivol is not significant. Following low skewness the coefficient on ivol is negatively significant at 0.10. Following high skewness the coefficient is positive. The difference in means of 0.14 is significant at 0.01.

I next consider the 420 observations of lagged $SKEW_{bm}$ and $\beta_{iv}$ and run a regression using White (1980) heteroscedasticity robust errors as in equation (1.14).

$$\beta_{ivol,\tau} = \gamma_0 + \gamma_{SK} \cdot SKEW_{bm,\tau-year} + \epsilon_{\tau}$$ (1.14)

$SKEW_{bm}$ predicts the sign of the relationship between ivol and returns when controlling for book-to-market in the cross-sectional regressions and when controlling for both book-to-market and firm size. In unreported regressions, including the BW sentiment index as a control variable does not alter the magnitude or the significance of the $SKEW_{bm}$ coefficient.

1.4.6 Aggregate Predictability

Lastly, I explore whether $SKEW_{bm}$ predicts aggregate market returns. The evidence on several sentiment indicators is that they predict the equal-weighted market index better than value-weighted one (Baker and Wurgler 2007.) This is because small stocks are more affected by sentiment. The effect on large stocks might sometimes be contrary to general sentiment as funds flow to-and-from the speculative smaller stocks.

I investigate whether $SKEW_{bm}$ can forecast the excess returns of the CRSP equal-weighted portfolio over the risk-free rate. Predictive regressions for the log
excess return over periods of one to five years are run. The dependent variable is the level of $SKEW_{bm}$ at the end of the previous month. Standard errors are calculated using Newey-West with $q=K-1$.

The results in Table 1.12 show that $SKEW_{bm}$ does predict the equal-weighted excess return. The coefficient is significant even after five years. The R-squared and the t-stats increase until three years and then drop. One standard deviation change in $SKEW_{bm}$ predicts 5.4 percentage points change in annual excess return for the equal-weighted portfolio over the next twelve months. This prediction falls to 3.3 annual percentage points for the five-year horizon.

On the other hand Panel B of Table 1.12 shows that $SKEW_{bm}$ is negatively correlated to future long-term excess returns on the value-weighted CRSP portfolio. Coefficients are for the most part not significant.

1.5 Conclusion

In this chapter I identify the skewness of the cross-section of book-to-market ratios as an important state variable. I demonstrate that it captures facets of the data that in other papers are attributed to sentiment. The variable is correlated with real macroeconomic variables as well as known sentiment indicators. Specifically, in good times the cross-sectional distribution of book-to-market ratios shows a left tail. During bad times this distribution is close to symmetric.

These findings are in line with the theory of limits to arbitrage, which predicts that while most stocks might be ‘correctly’ priced, at certain times certain stocks might display extreme mispricing. I.e. mispricing is not a smooth continuum but rather is concentrated in a subset of difficult to arbitrage stocks. When investor sentiment is positive extremely mispriced stocks are expected to be over-priced,
and when investor sentiment is negative severely mispriced stocks are expected to be under-priced.

This chapter suggests that the cross-sectional relationship between expected returns and book-to-market ratios is not a simple linear one, but rather convex. At times the value premium is due to low performance of very low book-to-market firms, and at others it is due to good performance of very high book-to-market firms. Therefore, papers, such as Cohen, Polk and Voulteneaho (2003), that estimate the ‘average’ linear relationship over time between book-to-market ratios and future returns, might in fact be missing some of the effect of investor sentiment on the cross-section of book-to-market ratios.

Previous literature on sentiment identifies patterns of returns based on firm characteristics that are related to the difficulty in arbitrage or in valuation. I show that the cross-sectional skewness captures these return patterns. For one, it predicts the returns on the Small Minus Big portfolio. It does so even when other known predictors are included in the regression. It also predicts returns on portfolios based on age, volatility, profitability and dividend payment controlling for size.

A prediction of the theory of the limits to arbitrage is that firms for which costs of arbitrage are higher will be more prone to mispricing. Following the literature I consider idiosyncratic volatility as a measure of difficulty in arbitrage. After periods of high cross-sectional skewness the more volatile firms outperform the less volatile ones. But, subsequent to periods of low skewness the volatile firms significantly underperform the less volatile ones.
Figure 1.1: Cross-sectional skewness of book-to-market ratios, 1970-2005.

The figure shows the skewness of the cross-section of book-to-market ratios of all stocks (share code 10 or 11) in NYSE/AMEX/NASDAQ. Book equity data are taken from the COMPUSTAT quarterly series. They are matched to market data skipping a full calendar quarter.
Figure 1.2: Monthly returns for portfolios based on long-term past returns, 1971-2005.

Every month ten portfolios based on past returns (-60,-13) are formed. Decile breakpoints are based on NYSE firms. Monthly returns for equal-weighted portfolios are calculated. Portfolios and returns are from Ken French’s website. Months are then classified into three groups by one-year lagged $SKEW_{bm}$. They are the lowest quartile of $SKEW_{bm}$, inter-quartile range and highest quartile. The lines show average returns for each decile’s equal-weighted portfolio in every group.
Figure 1.3: Monthly returns for portfolios based on short-term momentum, 1971-2005.

Every month ten portfolios based on past returns (-12,-2) are formed. Decile breakpoints are
based on NYSE firms. Monthly returns for equal-weighted portfolios are calculated. Portfolios
and returns are from Ken French’s website. Months are then classified into three groups by
one-year lagged $SKEWbm$. They are the lowest quartile of $SKEWbm$, inter-quartile range
and highest quartile. The lines show average returns for each decile’s equal-weighted portfolio
in every group.
Figure 1.4: Monthly returns for portfolios based on idiosyncratic volatility, 1971-2005.

Every month idiosyncratic volatility (ivol) is calculated for each firm. It is defined as the standard deviation of residuals in the monthly regression of excess returns on excess market returns over the previous five years. Firms are then ranked by ivol, and equal-weighted portfolios are formed based on NYSE/AMEX/NASDAQ deciles. Months are classified into three groups by one-year lagged \( SKEWbm \). They are the lowest quartile of \( SKEWbm \), inter-quartile range and highest quartile. The lines show average returns in excess of the market for each decile in every group.
Table 1.1: Summary Statistics 1970-2004.

Panel A considers $SKEW_{bm}$, the cross-sectional skewness of book-to-market ratios, and other conditioning variables. It uses monthly observations between 1970-2004. DPR, TBL, DFY and TMS are described in the text. DEMP is log growth in Total Seasonally-Adjusted Nonfarm Employment. DIP is growth in Industrial Production. EBM is the equal-weighted average of book-to-market ratios of CRSP firms. Panel B considers data from Jeff Wurgler’s Web site. They are described in the text. The panel includes 35 observations at year-end 1970-2005.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>mean</th>
<th>SD</th>
<th>Correlation with $SKEW_{bm}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Monthly Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SKEW_{bm}$</td>
<td>420</td>
<td>-0.72</td>
<td>0.27</td>
<td>1.0</td>
</tr>
<tr>
<td>Log Dividend-Price Ratio of S&amp;P 500 (DPR)</td>
<td>420</td>
<td>-3.52</td>
<td>0.45</td>
<td>-0.14***</td>
</tr>
<tr>
<td>T-bill (TBL)</td>
<td>420</td>
<td>6.08</td>
<td>2.91</td>
<td>-0.30***</td>
</tr>
<tr>
<td>Default spread (DFY)</td>
<td>420</td>
<td>1.09</td>
<td>0.04</td>
<td>-0.04</td>
</tr>
<tr>
<td>Term spread (TMS)</td>
<td>420</td>
<td>1.94</td>
<td>1.51</td>
<td>0.04</td>
</tr>
<tr>
<td>Growth in Employment (DEMP)</td>
<td>420</td>
<td>0.15</td>
<td>0.21</td>
<td>-0.17***</td>
</tr>
<tr>
<td>Growth in Industrial Production (DIP)</td>
<td>420</td>
<td>0.22</td>
<td>0.73</td>
<td>-0.13***</td>
</tr>
<tr>
<td>Equal-weighted Book-to-Market (EBM)</td>
<td>420</td>
<td>0.97</td>
<td>0.32</td>
<td>0.50***</td>
</tr>
<tr>
<td>Panel B: End-of-year Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dividend Premium (PDND)</td>
<td>35</td>
<td>-7.06</td>
<td>15.88</td>
<td>0.54***</td>
</tr>
<tr>
<td>Number of IPOs (NIPO)</td>
<td>35</td>
<td>372.03</td>
<td>272.59</td>
<td>-0.57***</td>
</tr>
<tr>
<td>First day returns to IPOs (RIPO)</td>
<td>35</td>
<td>16.99</td>
<td>14.72</td>
<td>-0.17</td>
</tr>
<tr>
<td>NYSE Turnover (TURN)</td>
<td>35</td>
<td>0.12</td>
<td>0.18</td>
<td>-0.32*</td>
</tr>
<tr>
<td>Close end fund discount (CEFD)</td>
<td>35</td>
<td>9.50</td>
<td>6.97</td>
<td>0.27</td>
</tr>
<tr>
<td>Share of equity in issuance (S)</td>
<td>35</td>
<td>0.19</td>
<td>0.09</td>
<td>-0.19</td>
</tr>
<tr>
<td>First Principal (SENT)</td>
<td>35</td>
<td>0.12</td>
<td>0.98</td>
<td>-0.35**</td>
</tr>
</tbody>
</table>
Table 1.2: Future Returns by SKEWbm and Book-to-Market Deciles.

The table reports monthly returns of portfolios by book-to-market deciles. Every month during 1971-2005 equal-weighted portfolios according to NYSE book-to-market decile breakpoints are formed. Average portfolio returns are grouped by lagged one-year SKEWbm in the following way: the sample of 420 months is broken down to low skewness periods defined as the lowest quartile, mid-skewness periods, which fall in the interquartile range, and the highest quartile of skewness periods. The last two columns compare the returns of portfolios 5 and 1, and portfolios 10 and 5.

<table>
<thead>
<tr>
<th>Deciles</th>
<th>SKEWbm</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>5-1</th>
<th>10-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOW</td>
<td>0.19</td>
<td>0.85</td>
<td>1.00</td>
<td>1.24</td>
<td>1.33</td>
<td>1.41</td>
<td>1.35</td>
<td>1.53</td>
<td>1.37</td>
<td></td>
<td></td>
<td>1.26</td>
<td>-0.08</td>
</tr>
<tr>
<td>MID</td>
<td>0.61</td>
<td>1.02</td>
<td>1.19</td>
<td>1.37</td>
<td>1.51</td>
<td>1.60</td>
<td>1.68</td>
<td>1.80</td>
<td>2.15</td>
<td>0.74</td>
<td>0.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIGH</td>
<td>1.66</td>
<td>1.55</td>
<td>1.62</td>
<td>1.52</td>
<td>1.43</td>
<td>1.51</td>
<td>1.70</td>
<td>1.64</td>
<td>1.93</td>
<td>2.85</td>
<td>-0.23</td>
<td>1.42</td>
<td></td>
</tr>
</tbody>
</table>

36
Table 1.3: Predicting Returns of Book-to-Market Portfolios.

Every month during 1971-2005 three equal-weighted portfolios are formed based on NYSE decile breakpoints for book-to-market ratios. High, Mid and Low are correspondingly the top three, middle four and bottom three deciles. Two long-short portfolios are formed: High minus Mid and Mid minus Low. The table reports results for regressions of these portfolio returns on lagged $SKEWbm$, the Fama-French factors (RMRF, SMB and HML) and a momentum factor (UMD). Panel A considers the High minus Neutral strategy:

$$ R_{H,t} - R_{M,t} = \beta_0 + \beta_{SK} SKEWbm_{t-\text{year}} + \gamma \cdot Controls_t + \epsilon_t $$

Panel B considers the Neutral minus Low strategy:

$$ R_{M,t} - R_{L,t} = \beta_0 + \beta_{SK} SKEWbm_{t-\text{year}} + \gamma \cdot Controls_t + \epsilon_t $$

The $SKEWbm$ variable is standardized to have a mean of zero and a standard deviation of one. In each of the panels the regression (1) is a univariate regression. Regression (2) and (4) include controls for the four factors. Regressions (3) and (4) use the residuals of the regression of $SKEWbm$ versus Baker and Wurgler’s monthly sentiment variable from Wurgler’s website. ** and *** denote significance at the 1% and 5% levels, respectively.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.39***</td>
<td>0.35***</td>
<td>0.39***</td>
<td>0.35***</td>
</tr>
<tr>
<td></td>
<td>(4.96)</td>
<td>(4.87)</td>
<td>(4.93)</td>
<td>(4.91)</td>
</tr>
<tr>
<td>$SKEWbm$</td>
<td>0.27***</td>
<td>0.25***</td>
<td>0.29***</td>
<td>0.26***</td>
</tr>
<tr>
<td></td>
<td>(3.56)</td>
<td>(3.89)</td>
<td>(3.51)</td>
<td>(3.61)</td>
</tr>
<tr>
<td>Four factors</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orthogonal to BW</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj-R²</td>
<td>0.02</td>
<td>0.38</td>
<td>0.02</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Table continued on next page ...
Panel B: Neutral minus Low

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.49***</td>
<td>0.33***</td>
<td>0.48***</td>
<td>0.33***</td>
</tr>
<tr>
<td></td>
<td>(3.76)</td>
<td>(3.95)</td>
<td>(3.72)</td>
<td>(3.97)</td>
</tr>
<tr>
<td>$SKEW_{bm}$</td>
<td>-0.30**</td>
<td>-0.14**</td>
<td>-0.21</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>(2.42)</td>
<td>(2.80)</td>
<td>(1.52)</td>
<td>(0.78)</td>
</tr>
<tr>
<td>Four factors</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orthogonal to BW</td>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Adj-R(^2)</td>
<td>0.01</td>
<td>0.72</td>
<td>0.00</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Table continued from previous page.
Table 1.4: Cross-Sectional Regressions for Size and Book-to-Market Conditional on $SKEW_{bm}$.

The table reports Fama-MacBeth regressions where the assets are 25 portfolios based on size and BE/ME. Each month the following regression is run:

\[ R_{i,t} - R_{f,t} = \beta_{0,t} + \beta_{ME,t} \log(ME)_{i,t-1} + \beta_{BM,t} \log(B/M)_{i,t-1} + \epsilon_{i,t}. \]

Panel A considers the time-series average coefficients for the full sample and then a split on the one-year lagged $SKEW_{bm}$ measure: lowest quartile, inter-quartile and highest quartile over the period 1970-2004. The last line considers the difference in means between high and low periods. Panel B considers a time-series regression for the cross-sectional coefficients on lagged $SKEW_{bm}$. The $SKEW_{bm}$ variable is standardized to have a mean of zero and a standard deviation of one. Regressions marked ‘Orthogonal to BW’ use the residuals of the regression of $SKEW_{bm}$ versus Baker and Wurgler’s monthly sentiment variable from Wurgler’s website. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

### Panel A: Fama-MacBeth means and t-stats for full sample and subsamples

<table>
<thead>
<tr>
<th>Period</th>
<th>N</th>
<th>$\beta_0$</th>
<th>$\beta_{ME}$</th>
<th>$\beta_{BM}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Sample</td>
<td>420</td>
<td>1.22***</td>
<td>-0.06</td>
<td>0.32***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.80)</td>
<td>(1.19)</td>
<td>(3.14)</td>
</tr>
<tr>
<td>Low - Q25</td>
<td>105</td>
<td>0.00</td>
<td>0.14</td>
<td>0.45**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.00)</td>
<td>(1.49)</td>
<td>(1.95)</td>
</tr>
<tr>
<td>Q25-Q75</td>
<td>210</td>
<td>1.16*</td>
<td>-0.04</td>
<td>0.44***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.89)</td>
<td>(0.56)</td>
<td>(3.08)</td>
</tr>
<tr>
<td>High - Q75</td>
<td>105</td>
<td>2.56***</td>
<td>-0.29***</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.81)</td>
<td>(3.25)</td>
<td>(0.22)</td>
</tr>
<tr>
<td>High minus Low</td>
<td>2.56**</td>
<td>-0.43***</td>
<td>-0.49*</td>
<td></td>
</tr>
</tbody>
</table>

Table continued on next page ...
Panel B: $\beta_{s,\tau} = \gamma_0 + \gamma_{SK} \cdot SKEWbm_{\tau-year}$

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
<th>ME</th>
<th>ME</th>
<th>B/M</th>
<th>B/M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.19***</td>
<td>1.18***</td>
<td>-0.05</td>
<td>-0.05</td>
<td>0.33***</td>
<td>0.33***</td>
</tr>
<tr>
<td></td>
<td>(2.47)</td>
<td>(2.46)</td>
<td>(1.06)</td>
<td>(1.05)</td>
<td>(2.98)</td>
<td>(2.97)</td>
</tr>
<tr>
<td>$SKEWbm$</td>
<td>1.28***</td>
<td>1.35***</td>
<td>-0.18***</td>
<td>-0.19***</td>
<td>-0.20**</td>
<td>-0.17*</td>
</tr>
<tr>
<td></td>
<td>(2.57)</td>
<td>(2.49)</td>
<td>(4.23)</td>
<td>(4.14)</td>
<td>(2.00)</td>
<td>(1.63)</td>
</tr>
<tr>
<td>Orthogonal to BW</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj-$R^2$</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
<td>0.01</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table continued from previous page.
Table 1.5: Cross-Sectional Regressions for Fama-French Factor Premia Conditional on $SKEW_{bm}$.

The table reports Fama-MacBeth regressions where the assets are 25 portfolios based on size and BE/ME. Factor betas are calculated using 5-year rolling regressions, and then each month, the following cross-sectional regression is run:

$$R_{i,t} - R_{f,t} = \lambda_{0,t} + \lambda_{R,t}\beta_{R,t-1} + \lambda_{S,t}\beta_{S,t-1} + \lambda_{H,t}\beta_{H,t-1} + \epsilon_{i,t}$$

The first line considers the full sample mean and Fama-Macbeth t-stats, and then a split by the lagged $SKEW_{bm}$ measure: lowest quartile, inter-quartile and highest quartile over the period 1970-2004. The last line considers the difference in means between high and low periods. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

Panel A: Fama-MacBeth means and t-stats for full sample and subsamples

<table>
<thead>
<tr>
<th>Period</th>
<th>Months</th>
<th>$\beta_0$</th>
<th>$\lambda_{RMRF}$</th>
<th>$\lambda_{SMB}$</th>
<th>$\lambda_{HML}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Sample</td>
<td>420</td>
<td>1.30***</td>
<td>-0.77***</td>
<td>0.17</td>
<td>0.46***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.93)</td>
<td>(2.69)</td>
<td>(1.05)</td>
<td>(3.02)</td>
</tr>
<tr>
<td>Low - Q25</td>
<td>105</td>
<td>1.62***</td>
<td>-0.95*</td>
<td>-0.30</td>
<td>0.64**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.18)</td>
<td>(1.67)</td>
<td>(1.08)</td>
<td>(1.99)</td>
</tr>
<tr>
<td>Q25-Q75</td>
<td>210</td>
<td>1.45***</td>
<td>-0.84**</td>
<td>0.05</td>
<td>0.56***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.77)</td>
<td>(2.03)</td>
<td>(0.21)</td>
<td>(2.59)</td>
</tr>
<tr>
<td>High - Q75</td>
<td>105</td>
<td>0.67</td>
<td>-0.46</td>
<td>0.86***</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.33)</td>
<td>(0.83)</td>
<td>(2.80)</td>
<td>(0.23)</td>
</tr>
<tr>
<td>High minus Low</td>
<td>-0.95</td>
<td>0.49</td>
<td>1.16***</td>
<td>-0.58</td>
<td></td>
</tr>
</tbody>
</table>

41
Table 1.6: Time Series Regressions of SMB Portfolio, 1971-2005.

The table reports regressions for predicting one-month return of the Small-Minus-Big as defined by Fama and French (1993). The regressors are one-year lagged \( \text{SKEW}_{\text{bm}} \) and controls, which are a dummy for January, one-year lagged equal-weighted book-to-market ratio (EBM), term spread (TMS), T-bill (TBL), default spread (DFY). Bond variables are lagged one quarter. Macro variables concurrent with \( \text{SKEW}_{\text{bm}} \) are growth in employment, growth in industrial production and NBER recession dummy. The \( \text{SKEW}_{\text{bm}} \) variable is standardized to have a mean of zero and a standard deviation of one. Regressions marked ‘Orthogonal to BW’ use the residuals of the regression of \( \text{SKEW}_{\text{bm}} \) versus Baker and Wurgler’s monthly sentiment variable from Wurgler’s website. \(*\), \(*\) and \(*\) denote significance at the 1%, 5% and 10% levels, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>0.19</th>
<th>0.02</th>
<th>0.02</th>
<th>-0.13</th>
<th>-0.10</th>
<th>0.24</th>
<th>-0.75</th>
<th>-0.47</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(1.16)</td>
<td>(0.11)</td>
<td>(0.09)</td>
<td>(0.53)</td>
<td>(0.43)</td>
<td>(0.38)</td>
<td>(1.03)</td>
<td>(0.72)</td>
</tr>
<tr>
<td>SKEWbm</td>
<td>0.49***</td>
<td>0.49***</td>
<td>0.52***</td>
<td>0.40**</td>
<td>0.42**</td>
<td>0.43*</td>
<td>0.41**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.45)</td>
<td>(3.47)</td>
<td>(3.44)</td>
<td>(2.09)</td>
<td>(2.32)</td>
<td>(1.78)</td>
<td>(1.94)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JAN</td>
<td>2.02***</td>
<td>2.01***</td>
<td>1.97***</td>
<td>1.97***</td>
<td>2.04***</td>
<td>1.97***</td>
<td>1.97***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.25)</td>
<td>(3.23)</td>
<td>(3.27)</td>
<td>(3.27)</td>
<td>(3.34)</td>
<td>(3.23)</td>
<td>(3.24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EBM</td>
<td>0.49</td>
<td>0.72</td>
<td>1.18**</td>
<td>0.19</td>
<td>0.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.71)</td>
<td>(1.19)</td>
<td>(2.08)</td>
<td>(0.16)</td>
<td>(0.78)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMS</td>
<td>-0.12</td>
<td>-0.05</td>
<td>-0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.81)</td>
<td>(0.30)</td>
<td>(0.52)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TBL</td>
<td>-0.14*</td>
<td>-0.07</td>
<td>-0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.88)</td>
<td>(0.77)</td>
<td>(1.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DFY</td>
<td>0.86*</td>
<td>0.99*</td>
<td>0.92*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.83)</td>
<td>(1.89)</td>
<td>(1.75)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macro vars</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ortho. to BW</td>
<td>+</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj-(R^2)</td>
<td>0.02</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Table 1.7: Time Series Regressions for Size by Book-to-Market Portfolios and for Size by Past-Performance Portfolios, 1971-2005.

The table reports fifteen regressions. One-year lagged $SKEW_{bm}$ is the regressor in all of them. It has been standardized to have a mean of zero and a standard deviation of one. In Panel A the dependent variable is one month return for a portfolio of smallest quintile of firms in a given book-to-market quintile minus the return on the largest quintile of firms in the same book-to-market quintile. In Panel B the dependent variable is one month return for a portfolio of smallest quintile of firms in a given long-term reversal quintile minus the return on the largest quintile of firms in the same long-term reversal quintile. In Panel C the dependent variable is one month return for a portfolio of smallest quintile of firms in a given short-term momentum quintile minus the return on the largest quintile of firms in the same short-term momentum quintile. The monthly size, book-to-market, prior (13-60) return, and prior (12-2) return breakpoints are NYSE quintiles. Portfolio allocations and returns are from Ken French Web site. *** and ** denote significance at the 1% level and 5% level respectively.

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Intercept</th>
<th>$SKEW_{bm_{t-year}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1-S5/B1:</td>
<td>-0.33</td>
<td>0.99***</td>
</tr>
<tr>
<td></td>
<td>(1.04)</td>
<td>(3.21)</td>
</tr>
<tr>
<td>S1-S5/B2:</td>
<td>0.20</td>
<td>0.70***</td>
</tr>
<tr>
<td></td>
<td>(0.75)</td>
<td>(3.16)</td>
</tr>
<tr>
<td>S1-S5/B3:</td>
<td>0.31</td>
<td>0.56***</td>
</tr>
<tr>
<td></td>
<td>(1.34)</td>
<td>(3.03)</td>
</tr>
<tr>
<td>S1-S5/B4:</td>
<td>0.44**</td>
<td>0.76***</td>
</tr>
<tr>
<td></td>
<td>(2.02)</td>
<td>(3.91)</td>
</tr>
<tr>
<td>S1-S5/B5:</td>
<td>0.50**</td>
<td>0.95***</td>
</tr>
<tr>
<td></td>
<td>(2.20)</td>
<td>(3.90)</td>
</tr>
</tbody>
</table>

Table continued on next page ...
Panel B: Small Minus Big in Long-Term Reversal Quintiles

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Intercept</th>
<th>$SKEW_{bm_{t-year}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1-S5/R1:</td>
<td>0.07</td>
<td>1.12***</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td>(4.04)</td>
</tr>
<tr>
<td>S1-S5/R2:</td>
<td>0.19</td>
<td>0.47**</td>
</tr>
<tr>
<td></td>
<td>(0.93)</td>
<td>(2.41)</td>
</tr>
<tr>
<td>S1-S5/R3:</td>
<td>0.37**</td>
<td>0.60***</td>
</tr>
<tr>
<td></td>
<td>(2.07)</td>
<td>(3.70)</td>
</tr>
<tr>
<td>S1-S5/R4:</td>
<td>0.29</td>
<td>0.52***</td>
</tr>
<tr>
<td></td>
<td>(1.48)</td>
<td>(2.72)</td>
</tr>
<tr>
<td>S1-S5/R5:</td>
<td>-0.02</td>
<td>0.52**</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(2.21)</td>
</tr>
</tbody>
</table>

Panel C: Small Minus Big in Short-Term Momentum Quintiles

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Intercept</th>
<th>$SKEW_{bm_{t-year}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1-S5/R1:</td>
<td>0.46</td>
<td>1.47***</td>
</tr>
<tr>
<td></td>
<td>(1.22)</td>
<td>(4.58)</td>
</tr>
<tr>
<td>S1-S5/R2:</td>
<td>0.40</td>
<td>0.89***</td>
</tr>
<tr>
<td></td>
<td>(1.63)</td>
<td>(4.32)</td>
</tr>
<tr>
<td>S1-S5/R3:</td>
<td>0.60***</td>
<td>0.75***</td>
</tr>
<tr>
<td></td>
<td>(2.63)</td>
<td>(3.91)</td>
</tr>
<tr>
<td>S1-S5/R4:</td>
<td>0.61***</td>
<td>0.77***</td>
</tr>
<tr>
<td></td>
<td>(2.70)</td>
<td>(4.05)</td>
</tr>
<tr>
<td>S1-S5/R5:</td>
<td>0.69***</td>
<td>0.68***</td>
</tr>
<tr>
<td></td>
<td>(3.04)</td>
<td>(3.17)</td>
</tr>
</tbody>
</table>

Table continued from previous page.
Table 1.8: Long Horizon Time Series Regressions for SMB.
Panel A considers regressions for predicting log(SMB) returns over periods of 1-5 years using 1-month lagged $SKEWbm$. $SKEWbm$ is standardized to have mean zero and standard deviation of one. Panel B controls for other predictors for the 12-month regression. It includes $EWbm$, the equal-weighted log book-to-market of firms in CRSP. The BW sentiment index is a first principal component of six market-based sentiment indicators. The components are also considered separately: NIPO, number of IPOs in the last year; PDND, log difference of the average market-to-book ratio of dividend payers and nonpayers; CEFD, average difference between the net asset values of closed-end stock fund shares and their market value; $S$, gross equity issuance divided by gross equity plus gross long-term debt issuance; RIPO, average first day returns on IPOs; TURN, detrended ratio of share volume to shares listed on NYSE. Data are taken from Wurgler’s Web site. ***, ** and * denotes significance at the 1%, 5% and 10% levels respectively.

<table>
<thead>
<tr>
<th>SMB Portfolio (log returns)</th>
<th>Panel A: $r_{t,t+K} = \beta_0 + \beta_{SK} \cdot SKEWbm_{t-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast Horizon</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(1.31)</td>
</tr>
<tr>
<td>$SKEWbm$</td>
<td>0.06***</td>
</tr>
<tr>
<td></td>
<td>(5.62)</td>
</tr>
<tr>
<td>Adj $R^2$</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Table continued on next page ...
Panel B: \( r_{t,t+12} = \beta_0 + \beta_{SK} \cdot SKEW_{bm_{t-1}} + \beta_j \cdot Control_{j,t-1} \)

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>( SKEW_{bm} )</th>
<th>EWbm</th>
<th>BW Index (( \beta \times 100 ))</th>
<th>NIPO (( \beta \times 100 ))</th>
<th>PDND (( \beta \times 100 ))</th>
<th>CEFD (( \beta \times 100 ))</th>
<th>S</th>
<th>RIPO</th>
<th>TURN</th>
<th>Adj ( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.03**</td>
<td>0.03***</td>
<td>0.16***</td>
<td>-0.18</td>
<td>-0.17***</td>
<td>-0.15</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>-0.04</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>(2.40)</td>
<td>(2.50)</td>
<td>(3.11)</td>
<td>(0.17)</td>
<td>(3.21)</td>
<td>(1.54)</td>
<td>(0.07)</td>
<td>(0.14)</td>
<td>(0.40)</td>
<td>(1.07)</td>
<td>(0.21)</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.05***</td>
<td>(3.58)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.07***</td>
<td>0.03***</td>
<td>(2.48)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>0.07***</td>
<td>(6.76)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.06***</td>
<td>(5.75)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.11***</td>
<td>0.05***</td>
<td>(4.20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table continued from previous page.
Table 1.9: Cross-Sectional Regressions on Firm Characteristics Conditional on \textit{SKEWbm}.

The table reports the coefficients of a second stage regression. The first stage is a monthly cross-sectional regression of the form

\[ R_{et} = \beta_{0,t} + \beta_{ch,t} \cdot \text{Char}_{i,t-1} + \beta_{BM,t} \cdot \log(B/M)_{i,t-1} \]

or

\[ R_{et} = \beta_{0,t} + \beta_{ch,t} \cdot \text{Char}_{i,t-1} + \beta_{BM,t} \cdot \log(B/M)_{i,t-1} + \beta_{ME,t} \cdot \log(ME)_{i,t-1} \]

for columns with \(^\dagger\). \text{Char} is whether the firm paid dividends or was profitable in the previous fiscal year, its age (years in CRSP) or R&D / Assets in the last fiscal year. The sample includes monthly returns between 1971-2005. The second stage regression is

\[ \beta_{ch,\tau} = \gamma_0 + \gamma_{SK} \cdot \text{SKEWbm}_{\tau - \text{year}} \]

\(*\ast\ast\ast\) and \(*\ast\ast\) and \(*\ast\) denote significance at the 1%, 5% and 10% levels, respectively. \textit{SKEWbm} has been standardized to have a mean of zero and a standard deviation of one.

<table>
<thead>
<tr>
<th></th>
<th>Pay Div</th>
<th>Pay Div(^\dagger)</th>
<th>Profit+</th>
<th>Profit+ (^\dagger)</th>
<th>Age</th>
<th>Age(^\dagger)</th>
<th>RD/A</th>
<th>RD/A(^\dagger)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.16</td>
<td>0.13</td>
<td>-0.09</td>
<td>0.10</td>
<td>-0.00</td>
<td>0.00(*)</td>
<td>4.50(*\ast\ast\ast)</td>
<td>3.86(*\ast\ast\ast)</td>
</tr>
<tr>
<td></td>
<td>(0.70)</td>
<td>(0.74)</td>
<td>(0.41)</td>
<td>(0.57)</td>
<td>(1.07)</td>
<td>(1.66)</td>
<td>(4.44)</td>
<td>(4.23)</td>
</tr>
<tr>
<td>(SKEWbm)</td>
<td>-0.70(*\ast\ast\ast)</td>
<td>-0.44(*\ast\ast)</td>
<td>-0.65(*\ast\ast\ast)</td>
<td>-0.42(*\ast)</td>
<td>-0.02(*\ast\ast\ast)</td>
<td>0.01</td>
<td>2.52(*\ast\ast\ast)</td>
<td>1.72(*\ast)</td>
</tr>
<tr>
<td></td>
<td>(2.95)</td>
<td>(2.02)</td>
<td>(2.54)</td>
<td>(1.79)</td>
<td>(3.54)</td>
<td>(1.36)</td>
<td>(2.52)</td>
<td>(1.82)</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Table 1.10: Time Series Regressions for Long-Short Portfolios.

Every month (1971-2005) two equal-weighted portfolios, *High* and *Low*, are formed based on NYSE breakpoints for given characteristics. They correspondingly include firms in the top three and bottom three deciles. The table reports results for regressions of long-high short-low portfolios on lagged *SKEWbm*: (1) by itself, (2) controlling for the Fama-French factors and a momentum factor, (3) controlling for the BW Sentiment Index, (4) Controlling for the Four Factors and the Sentiment Index.

\[
R_{H,t} - R_{L,t} = \beta_0 + \beta_{SK} SKEW_{bm,t-\text{year}} + \gamma \cdot \text{Controls}_t + \epsilon_t
\]

Firm characteristics are described in the text. Below are estimates of \(\beta_{SK}\) and their t-values. *SKEWbm* is standardized with mean zero and variance one. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-0.70***</td>
<td>-0.34***</td>
<td>-0.64***</td>
<td>-0.27**</td>
</tr>
<tr>
<td></td>
<td>(3.82)</td>
<td>(2.99)</td>
<td>(3.05)</td>
<td>(2.04)</td>
</tr>
<tr>
<td>Volatility</td>
<td>0.96***</td>
<td>0.48***</td>
<td>0.86**</td>
<td>0.34*</td>
</tr>
<tr>
<td></td>
<td>(2.98)</td>
<td>(3.05)</td>
<td>(2.28)</td>
<td>(1.91)</td>
</tr>
<tr>
<td>ROE</td>
<td>-0.14**</td>
<td>-0.09</td>
<td>-0.13*</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>(2.09)</td>
<td>(1.56)</td>
<td>(1.91)</td>
<td>(1.23)</td>
</tr>
<tr>
<td>E&gt;0</td>
<td>-0.76***</td>
<td>-0.42**</td>
<td>-0.67**</td>
<td>-0.31</td>
</tr>
<tr>
<td></td>
<td>(3.01)</td>
<td>(2.26)</td>
<td>(2.31)</td>
<td>(1.48)</td>
</tr>
<tr>
<td>D/BE</td>
<td>-0.25**</td>
<td>-0.10</td>
<td>-0.21*</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>(2.21)</td>
<td>(1.37)</td>
<td>(1.83)</td>
<td>(0.60)</td>
</tr>
<tr>
<td>D&gt;0</td>
<td>-0.79***</td>
<td>-0.37***</td>
<td>-0.67**</td>
<td>-0.23</td>
</tr>
<tr>
<td></td>
<td>(3.30)</td>
<td>(2.59)</td>
<td>(2.31)</td>
<td>(1.41)</td>
</tr>
<tr>
<td>PPE/A</td>
<td>-0.44**</td>
<td>-0.11</td>
<td>-0.45*</td>
<td>-0.12</td>
</tr>
<tr>
<td></td>
<td>(2.12)</td>
<td>(0.96)</td>
<td>(1.90)</td>
<td>(0.90)</td>
</tr>
<tr>
<td>RD/A</td>
<td>0.57**</td>
<td>0.11</td>
<td>0.54*</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>(2.41)</td>
<td>(0.68)</td>
<td>(1.91)</td>
<td>(0.57)</td>
</tr>
<tr>
<td>Controls</td>
<td>None</td>
<td>FF</td>
<td>BW</td>
<td>BW + FF</td>
</tr>
</tbody>
</table>

48
Table 1.11: Fama-MacBeth Regressions for Portfolios Based on Idiosyncratic Volatility.

Panel A reports summary statistics for Fama-MacBeth regressions where each month the following cross-sectional regression is run:

\[ R_{i,t} = \beta_{0,t} + \beta_{iv,t} \text{ivol}_{i,t-1} + \beta_{BM,t} \log(B/M_{i,t-1}) + \epsilon_{i,t} \]

First the full sample mean and t-stat are reported and then a split on the one-year lagged \( SKEW_{bm} \) measure. The bottom line reports the difference in means between Hi and Lo periods. Panel B considers a time-series regression of the cross-sectional coefficient on \( \text{ivol} \) on lagged \( SKEW_{bm} \). The first column considers a cross-sectional regression as in Panel A with only \( \text{ivol} \) and \( \log(B/M) \) as regressors. The second column also includes size in the cross-sectional regressions, and the third includes beta as well. The \( SKEW_{bm} \) variable is standardized to have a mean of zero and a standard deviation of one. \(*\star\star\) and \(*\) denote significance at the 1%, 5% and 10% levels, respectively.

<table>
<thead>
<tr>
<th>Period</th>
<th>N</th>
<th>(\beta_0)</th>
<th>(\beta_{iv})</th>
<th>(\beta_{BM})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Sample</td>
<td>420</td>
<td>0.86***</td>
<td>0.02</td>
<td>0.36***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.64)</td>
<td>(0.95)</td>
<td>(4.71)</td>
</tr>
<tr>
<td>Lo - Q25</td>
<td>105</td>
<td>1.33***</td>
<td>-0.06*</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.36)</td>
<td>(1.84)</td>
<td>(1.02)</td>
</tr>
<tr>
<td>Q25-Q75</td>
<td>210</td>
<td>0.93***</td>
<td>0.02</td>
<td>0.42***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.52)</td>
<td>(0.87)</td>
<td>(3.77)</td>
</tr>
<tr>
<td>Hi - Q75</td>
<td>105</td>
<td>0.26</td>
<td>0.08*</td>
<td>0.47***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.76)</td>
<td>(1.82)</td>
<td>(2.85)</td>
</tr>
<tr>
<td>High minus Low</td>
<td>-1.07*</td>
<td>0.14***</td>
<td>0.34</td>
<td></td>
</tr>
</tbody>
</table>

Table continued on next page ...
Panel B: $\beta_{\text{vol}, \tau} = \gamma_0 + \gamma_{SK} \cdot SKEWbm_{\tau - \text{year}}$

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.74)</td>
<td>(0.29)</td>
<td>(0.19)</td>
</tr>
<tr>
<td>$SKEWbm$</td>
<td>0.05***</td>
<td>0.04**</td>
<td>0.03**</td>
</tr>
<tr>
<td></td>
<td>(2.96)</td>
<td>(2.26)</td>
<td>(2.34)</td>
</tr>
<tr>
<td>Cross-sectional controls</td>
<td>B/M</td>
<td>B/M, ME</td>
<td>B/M, ME, $\beta$</td>
</tr>
<tr>
<td>Adj-$R^2$</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table continued from previous page.
Table 1.12: Long Horizon Time Series Regressions for CRSP Indexes, 1971-2005.

The table considers regressions for predicting log excess returns on the equal-weighted CRSP portfolio over periods of one to five years using $SKEW_{bm}$ at the end of the previous month. Panel B considers regressions for predicting log excess returns on the value-weighted CRSP portfolio. The $SKEW_{bm}$ variable is standardized to have a mean of zero and a standard deviation of one. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

Panel A: CRSP Equal-Weighted Portfolio (log returns)

\[ r_{t,t+K} = \beta_0 + \beta_{SK} \cdot SKEW_{bm_{t-1}} \]

<table>
<thead>
<tr>
<th>Forecast Horizon</th>
<th>Intercept</th>
<th>12</th>
<th>24</th>
<th>36</th>
<th>48</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.07***</td>
<td>0.14***</td>
<td>0.20***</td>
<td>0.28***</td>
<td>0.38***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.65)</td>
<td>(2.99)</td>
<td>(3.13)</td>
<td>(3.07)</td>
<td>(5.00)</td>
<td></td>
</tr>
<tr>
<td>$SKEW_{bm}$</td>
<td>0.05**</td>
<td>0.10***</td>
<td>0.14***</td>
<td>0.15***</td>
<td>0.16***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.25)</td>
<td>(3.56)</td>
<td>(3.92)</td>
<td>(2.55)</td>
<td>(2.65)</td>
<td></td>
</tr>
<tr>
<td>Adj $R^2$</td>
<td>0.06</td>
<td>0.13</td>
<td>0.18</td>
<td>0.15</td>
<td>0.14</td>
<td></td>
</tr>
</tbody>
</table>

Panel B: CRSP Value-Weighted Portfolio (log returns)

\[ r_{t,t+K} = \beta_0 + \beta_{SK} \cdot SKEW_{bm_{t-1}} \]

<table>
<thead>
<tr>
<th>Forecast Horizon</th>
<th>Intercept</th>
<th>12</th>
<th>24</th>
<th>36</th>
<th>48</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.05**</td>
<td>0.09**</td>
<td>0.13**</td>
<td>0.17***</td>
<td>0.23***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.23)</td>
<td>(2.28)</td>
<td>(2.25)</td>
<td>(2.40)</td>
<td>(2.80)</td>
<td></td>
</tr>
<tr>
<td>$SKEW_{bm}$</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.05</td>
<td>-0.07*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.91)</td>
<td>(0.87)</td>
<td>(0.73)</td>
<td>(1.36)</td>
<td>(1.88)</td>
<td></td>
</tr>
<tr>
<td>Adj $R^2$</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 2

Information, Expected Utility and Portfolio Choice

We study the consumption-investment problem of an agent with a constant relative risk aversion preference function, who possesses information about the future prospects of a stock. We also solve for the value of information to the agent in closed-form. We find that information can significantly alter consumption and asset allocation decisions. For reasonable parameter ranges, information increases consumption in the vicinity of 25%. Information can shift the portfolio weight on a stock from zero to around 70%. Thus, depending on the stock beta, the weight on the market portfolio can be considerably reduced with information, causing the appearance of under-diversification. The model indicates that stock holdings of informed agents are positively related to wealth, unrelated to systematic risk, and negatively related to idiosyncratic uncertainty. We also show that the dollar value of information to the agent depends linearly on his wealth and decreases with both the propensity to intermediate consumption and risk aversion.
2.1 Introduction

Information is an important feature of financial market settings. For example, a large industry comprised of both buy-side and sell-side financial analysts produces information signals, and the notion that company insiders are likely to have private information about their firms has long been recognized. A few key theoretical questions arise in this context: How does long-lasting private information alter the consumption-investment problem of an agent? How do parameters such as the propensity for intermediate consumption, the degree of risk aversion, stock beta, and the market risk premium affect the consumption-investment problem of an agent with long-lived information about a stock? What is the value of long-term information for a risk averse agent? What is the effect of wealth on the value of private information?

In this chapter, we study the value of long-lived private information in an analytic framework that takes into account wealth effects and can readily be calibrated to market data. Private information in this chapter is modeled as a signal about the value of the stock at a future time, and is received by an atomistic agent. We characterize the optimal portfolio strategy and consumption for the agent, who has a power utility function (i.e., a CRRA preference function) over intermediate consumption and terminal wealth and allocates his wealth across the stock, the market portfolio and a risk-free asset within a continuous time economy. We also consider the value of the information signal, defined as the difference in the utility equivalent of the agent with and without the information. Our primary goal is to compute and understand the impact of the signal on dynamic portfolio and consumption choice and utility gain. Our closed-form solution also facilitates the understanding of the effect of subjective discount rates, the investment horizon, market risk premium and volatility, and stock beta.
as well stock volatility on portfolio choice in the presence of private information.

While we are able to calculate the optimal portfolio holdings of the agent as well as the value of private information in a fairly general setting which includes the possibility of intermediate consumption, this tractability comes at a cost. First, we do not consider frictions such as taxes and bid-ask spreads. More importantly, we assume that the trading by the agent does not affect the market price.\(^1\) This is justified if the trading volume by the informed agent is very small compared to the total trading volume.\(^2\) If the full-fledged equilibrium could be solved, our calculation would correspond to the limiting case where the fraction of informed agents goes to zero, so that the agent becomes atomistic. Alternatively, the value calculated in this chapter can be viewed as an upper bound to the value of private information.\(^3\)

In our framework, the value of private information is proportional to the agent’s wealth under CRRA utility; therefore, the more wealthy the agent, the more valuable is the information in dollar terms. This result is consistent with the casual observation that private information is more valuable for large shareholders; thus they have a greater incentive to collect it. In most previous literature on asymmetric information, agents have CARA utility, so that there are no wealth effects.

Further, our analysis shows that the value depends on the propensity for in-

\(^1\)This assumption is in the spirit of the analysis of Kahl, Liu, and Longstaff (2003), which explores the cost of lockup periods within a continuous time setting.

\(^2\)There is evidence (Cornell and Sirri, 1992, Meulbroek, 1992, and Chakravarty and McConnell, 1999) which indicates that insider trades may not have as strong an effect on the market price as is suggested by strategic models of insider trading. While we are hesitant to use this literature as justification for our assumption that the informed agent is small enough to not affect the market price, the evidence does suggest that many insiders make trading decisions under the assumption that they are atomistic.

\(^3\)For our comparative statics on the value of private information, however, we require the interpretation that the computed value corresponds to that for an atomistic agent.
termediate consumption; informed agents with greater consumption propensities find private information to be less valuable. In addition, the age (or investment horizon) of the agent also has an effect on the value of information. A longer investment horizon trades off the benefit of increased opportunities for intermediate consumption against more uncertain, and highly discounted terminal wealth. If the propensity for intermediate consumption is sufficiently high, greater opportunities for intermediate consumption enhance the value of private information for younger agents.

We also find that less risk averse agents take a more aggressive position in the stock, which increases the value of private information. For agents with a low elasticity of intertemporal substitution (or high risk aversion), the propensity to consume negatively influences their expected initial holding in the risky asset. An agent with low risk aversion, however, wishes to save for the future, and is expected to hold more stock initially to consume relatively more in the future as the propensity to consume increases.

In the special case of log utility (where $\gamma$, the power of the CRRA utility function, equals unity), the consumption to wealth ratio for an informed equals that for an uninformed, so that information does not affect intermediate consumption. In general, the ratio of the consumption-to-wealth ratio for the informed over the uninformed depends on $\gamma$. When $\gamma > 1$, increased precision of private information causes the informed to consume more (as a fraction of wealth) relative to the uninformed; in particular, an informed agent will consume more than an uninformed agent. When $\gamma < 1$, the informed are more patient and greater precision causes them to consume less as a proportion of wealth (relative to the uninformed), in order to more fully exploit private information through time; in particular, an informed agent will consume less than an uninformed agent.
Our analysis is related to the literature on investment analysts. There is a large market for financial advice, and investment analysts are widely followed by the popular press. Extensive empirical studies have been performed on the analyst industry.\(^4\) The incentives of investment analysts to provide advice depend on the value of analyst advice to investors, because that value is linked to the compensation that analysts receive for providing the advice. Understanding of the value of private information and how that information alters portfolio holdings could potentially shed further light on the analyst industry, and this chapter provides a base upon which further analyses may be conducted on the subject.

This work is also relevant to the extensive literature which focuses on the incentives of insiders to hold stock in their company.\(^5\) One strand of this body of work links insider holdings to future price movements (e.g., Seyhun, 1986, 1991, Hadlock, 1998), and another considers the link between insider-manager ownership and measures of firm valuation (e.g., Jensen and Murphy, 1976, Demsetz and Lehn, 1985, Core and Guay, 1999, and McConnell and Servaes, 1990). However, an analysis of how a risk averse insider dynamically allocates wealth between different assets, including the equity of his own company, is an issue about which not much is known, and our study takes first things first by providing an initial analysis of the subject. Understanding this topic is important, because it also is key to comprehending the theoretical linkage between incentives and stock ownership. In particular, potentially richer insights could be shed in future work on that linkage if a primary incentive to hold company’s stock, namely, the possibility of superior information about the firm’s cash flows, is understood better.


\(^5\)From an empirical standpoint, Bluhshan (1989) finds that in the cross-section, insiders own about 23% the outstanding stock of companies on average.
within a theoretical setting.

Finally, this work is related to the literature on under-diversification. Surprisingly, we find that the holding in the individual stock may approach as high as 80% even for relatively low values of the precision. This offers a theoretical rationale for why corporate executives may not be as well-diversified as conventional theory would suggest. The chapter is also related to the literature on what may appear to be excessive holdings in private investment (Moskowitz and Vissing-Jorgensen, 2002), familiar stocks (Huberman, 2001) and the literature on home bias (Brennan and Cao, 1997, Kang and Stulz, 1997). In each of these cases, strong information about a company or an asset class’ performance prospects may cause portfolios to appear considerably under-diversified. The lack of diversification, as we show, can be a rational response to superior (positive) information about assets’ future prospects.\(^6\) We also show that changes in stock holdings are positively correlated with future expected returns on the risky asset. This evidence is consistent with the literature that relates insider and institutional holdings to future returns.\(^7\)

The value of private information, of course, has been studied extensively in earlier literature, but largely in the CARA (constant absolute risk aversion) setting with normally distributed asset values. Research on this topic dates from the seminal arguments of Hayek (1945) and Hirshleifer (1971); and the Grossman and Stiglitz (GS) (1980) model stimulated analytical research by providing a tractable closed-form solution to the expected utility of informed agents, as well as the equilibrium proportion of agents that choose to become informed.

\(^6\) The finding of high holdings of an individual stock applies in the case of a positive signal. Even on average, however, short-selling constraints could impede a symmetric negative position (see, e.g., Hong and Stein, 2003, Ofek and Richardson, 2003, and Lamont and Jones, 2002). Hence, portfolios of agents with private investors may appear to be under-diversified in the cross-section.

\(^7\) See Rozeff and Zaman (1988), Seyhun (1992), and Gompers and Metrick (2001).
This basic CARA-normal framework also has been used to analyze a number of important scenarios, for example, the buying and selling of information (Admati and Pfleiderer, 1987, 1990), multiple securities (Admati, 1985), market breakdowns (Bhattacharya and Spiegel, 1991), and diverse information (Verrecchia, 1982, Diamond and Verrecchia, 1981, Hellwig, 1980).

The GS setting and related papers have undoubtedly yielded numerous valuable insights. However, the specific combination of the utility function and the normal distribution that is imposed for tractability has restricted the generality of conclusions that can be drawn. There have been significant and valuable attempts towards calculating the value of information for more general utility functions; however, they have been conducted within static models, and by using approximations to the equilibrium – viz. Peress (2004). While these papers do consider the impact of wealth and private information on portfolio choice, how intermediate consumption and wealth effects influence the value of private information in a dynamic setting with general CRRA preferences is an issue which remains unaddressed in their work. General preferences considered within an intertemporal economy could potentially yield richer economic insights and also help relating the model to empirical quantities. For example, one feature of the basic Grossman and Stiglitz (1980) model is that the analysis is done in terms of price levels, not returns. Thus, returns are ratios of normally distributed variables, with undefined means and variances. As an empirical matter, however, returns and their moments are the key quantities of interest in a cross-sectional setting, and CRRA utility specifications allow the primitive to be returns rather than price levels.
than prices. Second, the Grossman and Stiglitz (1980) setting allows for the calculation of dollar holdings, not percentage portfolio holdings, but again in comparing securities and agents, the proportional holdings are of greater interest. Thus, potential extensions of our framework lend themselves more readily to calibrations with market data.

Previous research also has conducted rich dynamic analyses within the CARA-normal framework, viz. Wang (1994), Brown and Jennings (1989), and Grundy and McNichols (1989). Other than the difference in the preference structure, another distinction between this work and these other papers relates to the timing of information arrival. While the agent in this work receives a long-lived signal at the initial date, those in these other papers receive short-lived signals at multiple dates. In particular, in Wang (1993), the signal received is valid only over the next trading period, whereas in the models of Brown and Jennings (1989) and Grundy and McNichols (1989) the signal can be exploited for a maximum of two dates. While Vives (1995) presents a model where agents have information valid over many periods, he also uses the CARA framework and does not consider the value (i.e., expected utility) of the information signals.

Our model is related to that of Liu and Longstaff (2004). These authors postulate an arbitrage opportunity which converges to a value of zero at a terminal date. In effect, their model considers the case of an agent who has perfect foresight about the future value of the security. In contrast, we consider a generalized setting in which the informed agent has imperfect information about the future value of the stock, which requires us to explicitly model how the insider dynamically learns about future stock value from the stochastic evolution of the stock price.

The rest of the chapter is organized as follows. Section 2.2 describes the
basic structure of the model. Section 2.3 describes optimal investment and consumption with private information. Section 2.4 considers the economic gain from information. Section 2.5 concludes. All proofs appear in the appendix.

2.2 The Economic Setting

The informed agent has a finite investment horizon $T' < \infty$. We will assume that the agent has a power utility function over the intermediate consumption and the final period wealth

$$U = E_0 \left[ \int_0^{T'} \alpha' e^{-\rho t} \frac{c_t^{1-\gamma}}{1-\gamma} dt + e^{-\rho T'} \frac{W_{T'}^{1-\gamma}}{1-\gamma} \right], \quad (2.1)$$

where $c_t$ and $W_t$ represent consumption and wealth, respectively, at time $t$. The parameter $\gamma$ is a measure of risk aversion as well as an inverse measure of the elasticity of intertemporal substitution, while $\alpha'$ represents the agent’s propensity to consume at intermediate time-points. The quantity $T' - t$ may be viewed as the “age” of the agent at time $t$. We postulate that there are two risky assets: a market portfolio and a stock. The agent allocates wealth across three assets: his own risky stock, the market portfolio, and a riskless asset.

Since the agent is atomistic, he does not influence the market price. However, other agents trade the risky assets and determine prices. The market value of the market portfolio, $P_t$, follows the process

$$dP_t = P_t [(r + \mu) dt + \sigma_m dB_t],$$

whereas that of the individual stock, $S_t$, evolves according to

$$dS_t = S_t [(r + \beta \mu) dt + \beta \sigma_m dB_t + \sigma_s dZ_t],$$

where $r$ is the riskfree rate. We assume that $B_t$ and $Z_t$ are two independent standard Brownian motions. It can be seen that the diffusion processes for the
stock and the market portfolio are correlated through the common term involving $dB_t$, and $dZ_t$ represents stock-specific, or idiosyncratic risk. Further, $\beta$ describes the systematic risk of the stock.

Note that only the systematic risk $dB_t$ is priced, in the sense that it is associated with a risk premium $\mu$, while there is no risk premium associated with $dZ_t$. Because of this, the portfolio of an agent without information will consist only of the market portfolio and the riskless asset.

It follows from the Brownian motion specification that the stock price at time $T$ is

$$S_T = S_0 e^{\left(r + \beta \mu - \frac{1}{2} \left(\beta^2 \sigma^2_m + \sigma^2_s\right)\right) T + \beta \sigma_m B_T + \sigma_s Z_T}.$$ 

We assume that the agent receives a private signal about the diffusion process $Z_t$. Specifically, the agent observes a signal $L$ about $Z_T$,

$$L = Z_T + \sigma \epsilon,$$  \hspace{1cm} (2.2)

where $\epsilon$ is a standard normal random variable.

We now characterize the stochastic process for the evolution of the stock’s value from the standpoint of the agent with private information. Note that at time $t$, $dZ_t$ is a mean-zero normal random variable with variance $dt$. Equation (2.2) implies that

$$L = Z_t + dZ_t + (Z_T - Z_{t+dt}) + \sigma \epsilon.$$ 

Therefore, the above equation implies that $L - Z_t$ is a signal on $dZ_t$, with noise $(Z_T - Z_{t+dt}) + \sigma \epsilon$ which has a variance $T - (t + dt) + \sigma^2 \epsilon \approx T - t + \sigma^2 \epsilon$. Standard filtering theory involving normal random variables implies that

$$d\hat{Z}_t = dZ_t - \frac{(L - Z_t)}{T - t + \sigma^2 \epsilon} dt.$$ 

61
is a standard normal random variable conditional on \( L \) and \( Z_t \). Thus, the original Brownian motion is an Ornstein-Uhlenbeck process in the information set of the agent,

\[
dZ_t = \frac{(L - Z_t)}{T - t + \sigma_L^2} dt + d\hat{Z}_t,
\]

with \( \hat{Z}_t \) being the standard Brownian motion in the information set of the informed agent. While the above arguments are descriptive, formal derivations of this result are provided in the mathematics literature.\(^{10}\)

Let us define a process we term the “spread” as

\[
\Lambda_t \equiv L - Z_t.
\]

This process satisfies

\[
d\Lambda_t = \frac{-\Lambda_t}{T - t + \sigma^2} dt - d\hat{Z}_t.
\]

(2.4)

Note that the spread has a mean of 0 and a time-varying mean reversion coefficient of \( \frac{1}{T - t + \sigma^2} \). The mean reversion coefficient decreases deterministically with time \( t \) and the mean reversion is highest at \( t = T \).

The spread \( \Lambda_t \) can be expressed as a weighted average of past \( d\hat{Z}_t \) realizations,

\[
\Lambda_t = \frac{T - t + \sigma^2}{T + \sigma^2} L - \int_0^t \frac{T - t + \sigma^2}{T - u + \sigma^2} d\hat{Z}_u.
\]

To the informed agent, the evolution of the stock price is given by

\[
\frac{dS_t}{S_t} = \left( r + \beta\mu + \frac{\Lambda_t \sigma_s}{T - t + \sigma^2} \right) dt + \beta\sigma_m dB_t + \sigma_s d\hat{Z}_t,
\]

(2.5)

where the evolution of \( \Lambda \) is given by equation (2.4). It is clear from the above expression that the instantaneous expected return on the stock, conditional on the information signal, is directly related to \( \Lambda_t \).

\(^{10}\)See, for example, Jacod (1985) or Pikovsky and Karatzas (1996). The latter paper also solves the portfolio problem of an investor with logarithmic preferences, who maximizes the utility of terminal wealth with imperfect knowledge of a risky asset’s final payoff.
Since \( \Lambda_t \) is determined given the paths of \( dB_t \) and \( d\hat{Z}_t \) up to time \( t \), \( d\hat{Z}_t \) is determined by \( dS_t \) and \( dP_t \):

\[
d\hat{Z}_t = \frac{1}{\sigma_s} \left( \frac{dS_t}{S_t} - \beta \frac{dP_t}{P_t} \right) - \frac{\Lambda_t}{T - t + \sigma_t^2} dt.
\]

We adopt the standard stochastic control approach to solve the asset allocation problem of the informed agent. Let \( \phi_t \) and \( \phi_{tm}^m \) denote the time \( t \) proportional holdings in the stock and the market, respectively. The wealth dynamics are given by

\[
dW_t = W_t \left( r + \mu \phi_t^m + \left( \beta \mu + \frac{\Lambda_t \sigma_s}{T - t + \sigma_t^2} \phi_t \right) \right) dt - c_t dt
\]

\[
+ W_t(\phi_{tm}^m \sigma_m dB_t + \phi_t(\beta \sigma_m dB_t + \sigma_s d\hat{Z}_t)).
\]

Note that the expected evolution of the wealth of the individual depends on \( \Lambda \). This indicates that \( \Lambda \) is expected to play a key role in determining the individual’s portfolio holdings, and it is to this issue we will now turn.

Equation (2.1) can be written as

\[
U = \alpha^\gamma E_0 \left[ \int_0^T \alpha^\gamma e^{-\rho t} \frac{e^{\gamma t} \epsilon_t^{1-\gamma}}{1-\gamma} dt + e^{-\rho T} \frac{e^{\gamma T} W_{1-\gamma}^1}{1-\gamma} \right],
\]

where

\[
\alpha = \frac{1}{A} + \left( \frac{1}{\alpha} - \frac{1}{A} \right) e^{-A(T-T')},
\]

with

\[
A = \frac{\rho - (1 - \gamma) \left( r + \frac{\mu^2}{2\gamma \sigma_m^2} \right)}{\gamma}.
\]

Following Merton (1971), we define the indirect utility function \( J \) by

\[
J(W, \Lambda, t) = \max_{\phi_t^m, \phi_t^m} E_t[U].
\]

It is well known that the indirect utility function has the following form

\[
J(W, \Lambda, t) = e^{-\rho t} \frac{W_{1-\gamma}^1}{1-\gamma} f^\gamma(t, \Lambda).
\]

The appendix proves the following proposition.
Proposition 1 The function $f$ in the indirect utility function $J$ is given by

$$f(t, \Lambda; T) = \alpha \int_t^T e^{a(t; s, T) + \frac{1}{2} b(t; s, T) \Lambda^2} ds + e^{a(t; T, T) + \frac{1}{2} b(t; T, T) \Lambda^2},$$

where $a$ and $b$ are given by

$$\gamma a(t; s, T) = \left(-\rho + (1 - \gamma) \left( r + \frac{\mu^2}{2\gamma \sigma_m^2} \right) \right)(s - t) + \frac{1}{2} \gamma \ln \left( \frac{s - t}{\gamma (T - s + \sigma^2_\epsilon)} + 1 \right),$$

$$\gamma b(t; s, T) = \frac{1 - \gamma}{(T - t + \sigma^2_\epsilon)} \frac{s - t}{s - t + \gamma (T - s + \sigma^2_\epsilon)}.$$

Note that when $\sigma_\epsilon \to \infty$, $\gamma a = \left(-\rho + (1 - \gamma) \left( r + \frac{\mu^2}{2\gamma \sigma_m^2} \right) \right)(s - t)$ and $b = 0$; in this case, we recover Merton’s (1971) standard results, which are derived in the absence of private information. Also, the dependence of function $f$ on $(r, \rho, \mu, \sigma_m^2)$ is very similar to that in Merton. In fact, it can be shown when $\alpha = 0$, the dependence is identical. As such, the variation of optimal consumption and portfolio choice and the value function with $(r, \rho, \mu, \sigma_m^2)$ is isomorphic to that in Merton’s model. In the rest of the chapter, our main focus will be on the dependence of consumption, portfolio choice, and expected utility on the variables that characterize the information signal, namely, $\Lambda_t$ and $\sigma_\epsilon$.

One striking property of $a(t)$ and $b(t)$ is that they are finite and well-defined for $\gamma$ and $t$. Kim and Omberg (1996) show that if the stock return is predictable by way of an Ornstein-Uhlenbeck process, the functions $a(t)$ and $b(t)$ can be infinite for $\gamma < 1$ for a finite $t$. As such, the trading opportunity offered by such return dynamics can be so great that the value function is infinity. However, the mean-reversion coefficient in the Ornstein-Uhlenbeck process is constant, while the spread in our case has a time-varying mean-reversion coefficient that increases with $t$, as shown in equation (2.4). Even though the mean-reversion coefficient
is still bounded in our model, the fact that it increases with time reduces the stochastic investment opportunity and thus leads to a finite value function.

2.3 Optimal Investment and Consumption Policies with Information

In this section, we analyze how information affects the investor’s consumption and portfolio choice.

2.3.1 The General Case

Our next proposition derives the optimal consumption policy and the portfolio weights chosen by the informed agent.

Proposition 2 1. The optimal consumption is given by

\[ c_t^* = \alpha W f^{-1}. \]  

2. The optimal portfolio weights are given by

\[
\phi_t^* = \frac{1}{\sigma_s} \left( \frac{1}{T - t + \gamma \sigma_e^2} + \frac{\alpha \int_t^T (b(t; T, T) - b(t; s, T)) e^{a(t; s, T) + \frac{1}{2} b(t; s, T) \Lambda_t^2} ds}{\alpha \int_t^T e^{a(t; s, T) + \frac{1}{2} b(t; s, T) \Lambda_t^2} ds + e^{a(t; T, T) + \frac{1}{2} b(t; T, T) \Lambda_t^2}} \right) \Lambda_t, 
\]

and

\[
\phi_t^{m*} = \frac{\mu}{\gamma \sigma_m^2} - \beta \phi_t^*.
\]

Qualitatively, the optimal consumption rate is proportional to the wealth, as expected for a power utility maximizer. For a given \( \Lambda \), the more the agent invests today, the more the advantage he can take of the information and consume more later, which is the so-called substitution effect. On the other hand, with
private information, the agent in effect has more resources by being better able
to predict stock price movements, and thus may want to consume more in earlier
periods, which is the so-called wealth effect. When $\gamma < 1$, the substitution effect
dominates, the agent will invest more in the stock and thus consume less. On
the other hand, for an agent with $\gamma > 1$, the wealth effect dominates, so that
spreading consumption over the whole period is more important and the agent
will consume more. From the same intuition, the consumption rate increases
(decreases) with $\Lambda$ for $\gamma > 1$ ($\gamma < 1$).

We can also obtain an interpretation of the parameter $\alpha$ in the utility function
represented by (2.6). Specifically, note from (2.10) that
\[
\frac{\partial c^*}{\partial W} = \frac{\alpha}{f}.
\]
It can be seen from (2.9) that $\alpha/f$ is increasing in $\alpha$ (since $a$ and $b$ do not involve
$\alpha$), so that $\alpha$ is a measure of the propensity to consume at intermediate time
points. Later, we will see how $\alpha$ influences the holdings of the informed agents.

To compute the optimal consumption-to-wealth ratio and optimal portfolio
weights, we will assume the following benchmark case: $\Lambda = 0.5$, $\sigma_\epsilon = 30\%$, $T = 1$
year, $\gamma = 3$, $\alpha = 1$, $\sigma_s = 40\%$, $r = 4\%$, $\rho = 0.2$, $\mu = 6\%$, $\sigma_m = 15\%$, and
$\beta = 1$ (only needed for computing the portfolio weight of the market). As we
discussed earlier, the effects due to $r$, $\rho$, $\mu$, and $\sigma_m$ are small. $\alpha = 1$ is used in
most literature. $\gamma = 3$ is standard. We will assume $T = 1$. Note that uncertainty
in $\hat{Z}_1$ is 1, so $\sigma_\epsilon = 30\%$ implies a reduction in volatility of $1 - \sqrt{1 - 0.3^2} = 4.6\%$
, which is quite small. Also observe that the contribution to the expected return
due to $\Lambda$ is $\frac{\Lambda \sigma_\epsilon}{T - t + \sigma_\epsilon^2} = \frac{0.5 \times 0.4}{1 + 0.3^2} = 18\%$. The benchmark case is indicated by $\ast$ in all
the figures to follow.

For the benchmark case, computations show the agent will increase the optimal
consumption by 25% relative to the case of no information. This is a quite
significant increase. As a comparison, note that the consumption-to-wealth ratio is just a deterministic function of time for a CRRA agent without information, such an agent would increase his consumption by 25% only if his wealth is increased by 25%.

Figure 2.1 plots the consumption to wealth ratio as a function of $\Lambda$ for $\gamma = 3$ and $\gamma = 1/2$, with the rest of the parameters the same as the benchmark case. Note that the consumption is an even function $\Lambda_t$, thus the agent will increase the consumption irrespective whether it is a good news or bad news. This is due to the fact that if $\Lambda_t$, the agent will just short the stock. Even at $\Lambda = 0$, the informed agent still increases his consumption by 21% relative to the uninformed. This is due to the fact that there is information content in a signal with $\Lambda = 0$ relative to the case when there is no information at all.

Figure 2.2 plots the consumption to wealth ratio (at time 0) of an informed agent relative to that of an uninformed one as a function of the risk aversion $\gamma$. The parameter values used are otherwise the same as the benchmark case. The figure shows that the informed consume relatively less than the uninformed for $\gamma$ smaller than 1, while the reverse is true for $\gamma$ larger than 1. These results can be explained by noting that the parameter $\gamma$ in the utility function is inversely related to the elasticity of intertemporal substitution. For small $\gamma$, the agent has a stronger tendency to substitute intertemporally, and consequently the consumption-to-wealth ratio is low. The reverse is true for high $\gamma$. Note, however, that the ratio is not strictly increasing in $\gamma$. Very high $\gamma$ investors are too risk averse to make any use of the uncertain information and the consumption to wealth ratio of the informed asymptotes to that of the uninformed.

Figure 2.3 presents the same quantity as Figure 2.1 but as a function of the noisiness of private information ($\sigma_t^2$). The figure shows that for log utility ($\gamma = 1$),
the consumption to wealth ratio for an informed equals that for an uninformed. In this case, myopia dictates that the informed is only concerned about the one-step ahead investment opportunity. On the other hand, for $\gamma > 1$, the bigger the precision, the more the informed consume relative to the uninformed. In this case, the low elasticity of intertemporal substitution dictates that the more precise information will be employed to increase current consumption. For $\gamma < 1$, the informed are more patient and choose to consume less to exploit the more precise information later on in the trading process. Regardless of the risk aversion, as the precision of the information vanishes the consumption to wealth ratio of the informed goes to that of the uniformed.

We now turn to the portfolio weights chosen by the informed agent as determined in Proposition 2. Note that neither the $J$ function nor the optimal stock portfolio weight depends on $\beta$. This is because the effect of $\beta$ can be undone by taking an offsetting position in the market and the optimal combined exposure to the market risk is completely determined by the market volatility $\sigma_m$ and market risk premium $\mu$. Furthermore, the holding of the stock is determined completely by the information advantage $\Lambda_t$ and the idiosyncratic risk $\sigma_s$ (which represents the risk the agent has to bear to take advantage of private information).

On other hand, the market portfolio weight $\phi_{mt}^*$ depends on $\beta$ linearly, due to the fact that the market needs to off-set the market risk exposure in the stock position. Note that the dependence of the market portfolio weight on $\Lambda_t$ and $\sigma_s$ is opposite to that of the stock portfolio weight as shown in the above equation (assuming $\beta > 0$). So, we will mainly focus on discussions of the stock portfolio weight below. This weight is proportional to $\Lambda_t$; thus, the agent will hold more stock with a larger positive $\Lambda_t$ and short more stock with a larger negative $\Lambda$, as one might intuitively expect.
For the benchmark case, the optimal stock portfolio weight is 74%. In contrast, for the uninformed agent, the holding of the stock should be zero. The market portfolio weight in this case 15% whereas without information it is 89%. Thus information dramatically alters the agent’s portfolio. With such an order of magnitude, information effect can potentially be used to explain why investor hold undiversified portfolios.

In Figure 2.4, the initial (time 0) holding is plotted as a function of the propensity to consume, \( \alpha \). For highly risk averse informed agents, the propensity to consume negatively influences their holding in the risky asset. In this case, the agent wishes to hold less stock at time 0 and consume more if \( \alpha \) is large. An agent who is less risk averse than log utility, however, has a greater tendency to postpone consumption for the future, and in this case holds more stock to consume relatively more in the future as \( \alpha \) increases. Observe that an informed agent with low risk aversion (\( \gamma = 0.5 \)) initially chooses to invest more than 100% of his wealth in the stock. The intuition is that the agent takes a more aggressive position to consume more in the future when the risk aversion is low.

In Figure 2.5 we present the initial holdings in the three assets, the stock, the market, and the risk free asset as a function of the noisiness parameter \( \sigma^2 \). We find that as the information becomes more imprecise, the holding in the stock decreases, while the holdings in the market and the riskless asset increase. This finding is intuitive. It is noteworthy that the proportion allocated to the individual stock can approach as high as 70% even for moderate values of \( \sigma^2 \). The model is thus related to the literature on investing in the familiar (Huberman, 2001) as well as that on home bias (Brennan and Cao, 1997, Kang and Stulz, 1997). In each of these cases, strong positive information about a company or an asset class’ performance prospects may cause portfolios to appear consider-
ably underdiversified. The under-diversification, as we show, can be a rational response to superior (positive) information about assets’ future prospects.

Note that for high values of $\sigma^2_t$, the holding in the risky stock dips below the holding in both the market as well as the risk-free asset. Thus, insiders with highly imprecise information will place greater emphasis on diversification than on holdings in their own stock. A prediction of this part of the analysis is that for companies where good information is hard to come by, such as the high tech sector, will have better-diversified insiders.\(^{11}\) Also note that $\phi^* \to 0$ as $\sigma_t \to \infty$; this is expected since the signal becomes completely uninformative in this limit.

### 2.3.2 The Case of Logarithmic Utility

The expressions for consumption and portfolio weights can be further simplified in the case of logarithmic utility ($\gamma = 1$), as shown in the appendix (within the proof of Proposition 3 to follow). Under this preference structure, the investor’s utility can be written as

$$U = \lim_{\gamma \to 1} E_0 \left[ \int_0^T \alpha e^{-\rho t} \frac{c_t^{1-\gamma} - 1}{1 - \gamma} dt + e^{-\rho T} W_T^{1-\gamma} - 1 \right]$$

$$= E_0 \left[ \int_0^T \alpha e^{-\rho t} \ln(c_t) dt + e^{-\rho T} \ln(W_T) \right]$$

(2.13)

Further, in this setting, the consumption to wealth ratio does not depend on the signal and is given by

$$\frac{c}{W} = \frac{\alpha}{f_1}$$

where

$$f_1 = \frac{\alpha}{\rho} \left[ 1 + \left( \frac{\rho}{\alpha} - 1 \right) e^{-\rho(T-t)} \right]$$

\(^{11}\)One way to obtain diversification of private wealth is to go public. Our analysis thus suggests that insiders in companies with highly uncertain cash flows will be more prone to indulge in IPOs.
From the first order conditions, the portfolio holdings are given by

\[ \phi_{t}^{m*} = \frac{\mu}{\sigma_m^2} - \beta \phi_{t}^*, \]  
(2.14)

\[ \phi_t = \frac{1}{\sigma_s T - t + \sigma_s^2} \Lambda_t, \]  
(2.15)

As can be seen, the consumption-to-wealth ratio is a non-stochastic function of the various parameters that do not involve the information signal. The myopic behavior implied by logarithmic utility dictates that the agent ignore the long-term value of the private signal in designing his optimal consumption policy. From (2.15), we see, however, that the holdings of the risky stock depends directly on \( \Lambda \). From the definition of \( \Lambda_t \) in (2.3), it is evident that the expected long-run return on the stock, i.e., \( \ln(S_T/S_0) \), is correlated with the informed agent’s initial holding of the risky asset.\(^{12}\) This accords with the empirical literature (e.g., Seyhun, 1986, 1991, Hadlock, 1998), which indicates that insider holdings predict future stock returns.

### 2.4 Utility Gains from Private Information

In this section, we present a series of comparative statics results that build intuition on the economic impact of various parameters on the consumption and investment of the informed, as well as the value of private information. We define the value of private information as the ratio of certainty equivalents with and without the signal, for a given set of parameter values.

Let us define

\[ a_0 \equiv \left( -\rho + (1 - \gamma) \left( r + \frac{\mu^2}{2\gamma\sigma_m^2} \right) \right) (s - t) \]

\(^{12}\)Of course, this finding holds for the general CRRA case as well, but the expressions are more complicated.
and let $f_0(t; T)$ be the value of $f(t; \Lambda, T)$ that corresponds to $a = a_0$ and $b = 0$, note that function $f_0$ does not depend on $\Lambda$ because $b = 0$. Then, $J_0 = \frac{W^{1-\gamma}}{1-\gamma} f_0^\gamma$ is the indirect utility of an agent without information. We obtain the following proposition.

**Proposition 3** The value of private information at time $t$, defined as the ratio of the certainty equivalent with the information signal to that without the signal, and denoted by $R(t)$, is given by

$$R(t) = \left( \frac{f(t, \Lambda; T)}{f_0(t; T)} \right)^{\frac{1}{1-\gamma}}, \quad (2.16)$$

and is always greater than unity.

Note that the ex ante value of private information (before the signal is realized), which we denote $R_v$, is given by $R_v \equiv (E[R(t)^{1-\gamma}])^{\frac{1}{1-\gamma}}$.

### 2.4.1 Numerical Illustrations

For the benchmark case from the previous section, the ex ante value of information, i.e., $R_v$ (at time 0), is 1.56. If the signal is more precise, i.e. $\sigma_\epsilon$ is 0.10, the value jumps to 2.22. Thus the expected utility of an agent who has a wealth of $1$ million and receives a signal with $\sigma_\epsilon = 0.05$ is the same as that of an agent with about $2.2$ million in wealth but who receives no signal.\(^{13}\)

We present the ex post value of information (i.e., after realization of the private signal) in Figure 2.6 as a function of the time horizon. As can be seen, the value of information first increases, and then decreases in the time horizon. The intuition is that increasing the time horizon has two effects: there are more

---

\(^{13}\)For highly wealthy individuals such as Bill Gates, the assumption of an atomistic agent is less reasonable (unless perhaps the stock position is accumulated slowly and anonymously). In such cases, as we mentioned in the introduction, our computed quantity can be viewed as an upper bound to the value of information for such agents.
opportunities to trade, but it is also more risky to hold a position in the stock. Hence, for small values of $T$, the former effect dominates, whereas for large values of $T$, the latter feature takes over. The figure also indicates that information is more valuable for less risk averse agents. This is because agents with low risk aversion are able to take a more aggressive position in the stock.

In Figure 2.7 we plot the ex ante value of information (before realization of the signal) as a function of the consumption parameter $\alpha$, using the same parameters as before. As can be seen, the greater is the propensity to consume, the smaller is the value of information. In addition the value of information is greater for low risk aversion. The drop in the value of information as a function of $\alpha$ is steeper for the low risk aversion ($\gamma = 0.5$) case. In this case, the agent wishes to exploit private information by consuming relatively less and saving more for the future. A high $\alpha$ shifts relatively more consumption early on in the trading process and thus sharply reduces the ex ante value of information. The basic notion is that agents who wish to consume more at intermediate time points find long-term information to be less valuable.

### 2.4.2 A Specific Calibration

To get a feel for the (ex post) value of information given the realization of a specific information signal with a tangible empirical interpretation, we calibrate our model to the case of earnings surprises. Campbell, Lo, and MacKinlay (1997) show that the 21 day cumulative abnormal return in advance of a positive earnings surprise is 1.966%. The analogous number for a negative surprise is $-1.539\%$. Since the CAR can be viewed as the cumulative realization of $\sigma_s dZ_t$, the signal $L$ can be calibrated as the cumulative abnormal return scaled by $\sigma_s$, the idiosyncratic volatility of the stock. We choose $\sigma_s = 0.25$ based on the find-
ings of Campbell, Lettau, Malkiel and Xu (2004) about the idiosyncratic volatility of individual stocks. Then, for a positive surprise, \( L \) can be calibrated as \( 0.01966 \times 250 / (21 \times 0.25) = 0.9362 \), and the corresponding number for a negative surprise is \(-0.7329\). Considering current Treasury Bill yields, we choose an annual risk-free interest rate of 4%, and, based on Siegel (1998), an equity risk premium of 8%. Somewhat subjectively, we set \( \gamma = 3 \), which is within (but towards the lower end of) the range considered by Prescott and Mehra (1997). The subjective discount rate (\( \rho \)) and the propensity to consume (\( \alpha \)) are set to be 0.1 and 0.2, respectively. The \( \beta \) is 0.8 and \( T \) is 0.084 (21 days).

Table 2.1 reports the ex-post value of information for the cases of a good news surprise and of a bad news surprise. Even when the signal is relatively noisy, such as in the case where \( \sigma_\epsilon \) equals 20%, the value of information is 4.31 in the case of good news, implying that an agent with $1 million in wealth has the same utility from the signal as an agent with $4.31 million in wealth, but without the signal.

Figure 2.8 plots how calibrated holdings in the stock vary with the variance of \( \sigma_\epsilon \) in response to a positive or negative earnings signal. The weight on the stock can be very high for the case of a precise signals, and diminishes as the signal becomes less precise. The figure vividly demonstrates how private information can significantly impact the holdings of informed agents.

### 2.4.3 The Special Case of No Intermediate Consumption

Consider the case where \( \alpha = 0 \). In this case the ratio of the utility equivalent of the informed agent to that of the uninformed is given by

\[
R(t) = \frac{(T-t)_{\sigma^2} + 1}{(T-t)_{\sigma^2} + 1} \exp\left( \frac{(T-t)A_{t}^2}{2(T-t + \sigma^2)(T-t + \gamma \sigma^2)} \right).
\]
Note that \( R > 1 \) even if \( \Lambda = 0 \). Knowing that \( Z_T \) will equal \( Z_t \) is still more valuable than knowing nothing. The increased value is due to trading before \( T \). Even though \( \Lambda_t \) may be zero today, the future spread \( \Lambda \) may become non-zero and the informational advantage can thereby be exploited between \( t \) and \( T \). The value of information depends positively on \( \Lambda_t^2 \), which is intuitive.

Using the explicit expression above, we can verify the following. When there is no information, i.e., \( \sigma \epsilon \to \infty \), \( R(t) \to 1 \). In the special case where \( \sigma \epsilon = 0 \), the informed agent knows the stock price at time \( T \) precisely and the return of the stock follows a Brownian bridge process. As \( \sigma \epsilon \to 0 \), \( R(t) \to \infty \), implying that utility with information approaches infinity when the noise in information approaches 0. Liu and Longstaff (2004) study a version of this special case by modeling an arbitrage opportunity whose final value converges to zero. However, they obtain a finite utility level because, in their paper, there are margin requirements which place restrictions on the position the agent can take.

Note that, when \( \alpha = 0 \), with the exception of \( \gamma \), \( R \) only depends on the characteristics of the signal, such as time to revelation of information, \( T - t \), the signal precision, \( \sigma \epsilon \), and the spread \( \Lambda_t \). The ratio \( R \) does not depend on the interest rate, market risk premium, market volatility, stock beta, and stock volatility. One might expect otherwise because the riskless asset and the market portfolio are the alternative to investment in the stock. This happens because the information affects the indirect utility function \( J \) multiplicatively when \( \alpha = 0 \), as can be seen from Proposition 1. That is, the indirect utility function is the indirect utility for the no information case multiplied by a factor that is affected by private information.

It is also of interest to calculate the value of information at the time the
information is received (i.e., time 0). We have that
\[ \!
\begin{align*}
R(0) &= \frac{\left(\frac{1}{\sigma^2} + 1\right)^{\frac{1}{\gamma}}}{\left(\frac{1}{\gamma\sigma^2} + 1\right)^{\frac{1}{2(1-\gamma)}}} \exp\left(\frac{L^2}{2(1 + \sigma^2\epsilon)(1 + \gamma\sigma^2\epsilon)^{T}}\right).
\end{align*}
\]

As can be seen, the ratio of the utility equivalents is related to the square of the signal. Also, since only a fraction \( \phi^*_0 \) of the initial wealth is responsible for generating the differential in utility equivalents represented in \( R(0) \), the net proportional increase in the stock holding is
\[ \!
\begin{align*}
R_s(0) &= (R(0)-1)/\phi^*_0 = \frac{\left(\frac{1}{\sigma^2} + 1\right)^{\frac{1}{\gamma}}}{\left(\frac{1}{\gamma\sigma^2} + 1\right)^{\frac{1}{2(1-\gamma)}}} \exp\left(\frac{L^2}{2(1 + \sigma^2\epsilon)(1 + \gamma\sigma^2\epsilon)^{T}}\right) \frac{T(1 + \gamma\sigma^2\epsilon)\sigma_s}{L}.
\end{align*}
\]

We now present an expression for the ex-ante value of the information.

**Proposition 4** The ex-ante value of information at any given time \( t \) is
\[
R_v = \left( E \left[R(t)^{1-\gamma}\right] \right)^{\frac{1}{1-\gamma}} = \sqrt{1 + \frac{T - t}{\gamma\sigma^2\epsilon}}.
\]  
(2.17)

As can be seen, the ex ante value of information is always greater than 1. At any time \( t \), it is increasing in the ratio of the variance of the brownian motion of the stock return \( (T - t) \) over the variance of the signal \( (\sigma^2\epsilon) \) and decreasing in the risk aversion. Even if the signal noise has greater variance than the underlying stock the additional information is still valuable, because the signal helps reduce uncertainty about the stock’s terminal value.

**2.4.4 Differing Information and Investment Horizons**

Note that in our model, the agent has information about the stock value at time \( T \) but has an investment horizon until time \( T' > T \). From (2.6)-(2.8), this problem is equivalent to that of an investor who has an investment horizon of \( T \) and a
propensity to consume $\alpha' = \frac{1}{K_0}$, where $K_0$ is given by

$$K_0 = \frac{1}{A} + \left( \frac{1}{\alpha} - \frac{1}{A} \right) e^{-A(T'-T)},$$

and where

$$A = -\frac{-\rho + (1 - \gamma) \left( r + \frac{\mu^2}{2\gamma\sigma^2} \right)}{\gamma}.$$

The value of a longer investment horizon is a tradeoff between additional intermediate consumption and a farther out, and therefore riskier and highly discounted, terminal wealth. In Figure 2.9 we plot the ex ante value of information as a function of the investment horizon for various levels of risk aversion. In all the plotted cases the propensity to consume is large enough that the effect of additional opportunities to consume dominates, so that the ex ante value of information monotonically increases with the investment horizon.

To build more intuition in this case, Figure 2.10 shows the holdings in stock at time zero as a function of the investment horizon. Since in this case an opportunity for more intermediate consumption is more valuable than a terminal wealth that is nearer in the future, the consumption stream tends to become smoother with a longer investment horizon. Therefore, as can be seen in the figure, highly risk averse investors will increase their time zero holding in stock and postpone their consumption as their investment horizon increases. Conversely, the time zero holding in stock decreases for the low risk aversion individuals as they face a longer investment horizon. Myopic log-utility investors are of course indifferent to the investment horizon.

### 2.5 Conclusion

We analyze the consumption-investment problem of an agent with CRRA preferences in a continuous time setting. For tractability, we assume that the agent
is atomistic, which leads to an analytic expression for consumption, portfolio weights, and the ex ante value of private information for such an agent. Our analysis provides a link between the literature on dynamic portfolio choice, which principally uses power utility functions, and that on informed investment, which typically assumes CARA preferences or risk neutrality. In addition, our model allows the characterization of the value of information in terms of empirically measurable quantities. This allows a calibration of the model to real-world data, which has implications for academics (and policymakers) who are interested in the magnitude of gains from information in equity markets.

Our analysis indicates that information is worth more in dollar terms, the greater is the wealth of agents, unlike in the case of exponential preferences, in which instance the value of information is independent of wealth. Since we explicitly model the propensity to consume at intermediate time points, we are able to examine how consumption alters the value of private information. Thus, we find that informed agents who have greater propensities to consume at intermediate times find long-term private information to be less valuable since they are less able to fully obtain the long-term benefits of trading on such information.

We also show that information can have significant effects on consumption as well as asset allocation. For reasonable parameter ranges, information can increase the consumption in the vicinity of 25%. An investor holding in the stock may approach 75% with information compared to 0% were he not to have any information. Thus, agents with private information about an investment opportunity may appear to be substantially overinvested in that opportunity, which sheds light on the under-diversification phenomenon documented in various settings. We conduct a calibration exercise and find that even a noisy signal about a good or bad earnings surprise can significantly amplify expected utility.
Further, insider holdings in the risky asset are related to future expected returns on that stock, which is consistent with the analyses of Seyhun (1986, 1992), and Rozeff and Zaman (1988).

The model suggests that portfolio holdings are positively related to wealth, inversely related to idiosyncratic uncertainty, and unrelated to systematic risk. We note that there are aspects of our theoretical analysis that could be extended to other settings. First, adapting our framework explicitly to multiple, correlated assets would be interesting and allow for predictions about insider holdings in related stocks, possibly those in the same industry. Second, while this is a difficult analytical issue, a solution to the full rational expectations setting where the insider is not atomistic remains elusive. A search for such a solution is clearly a predominant part of the agenda for future work on the subject.

Appendix

Proof of Propositions 1-3: We follow Merton (1971) in defining the indirect utility function \( J \) by

\[
J(W, V, t) = \max_{c_t, \phi_t, \phi_t^m} E_t[U(W_T)]
\]

\[
dW_t = W_t \left( r + \mu \phi_t^m + \left( \beta \mu + \frac{\Lambda_t \sigma_s}{T - t + \sigma_e} \right) \phi_t \right) dt - c_t dt + W_t \left( \phi_t^m \sigma_m dB_t + \phi_t \left( \beta \sigma_m dB_t + \sigma_s d\hat{Z}_t \right) \right)
\]

\[
= W_t \left( r + \mu \phi_t^m + \beta \phi_t + \frac{\Lambda_t \sigma_s}{T - t + \sigma_e} \phi_t \right) dt - c_t dt + W_t \left( \phi_t^m \sigma_m dB_t + \phi_t \sigma_s d\hat{Z}_t \right)
\]

\[
= W_t \left( r + \mu \phi_t^m + \frac{\Lambda_t}{T - t + \sigma_e} \phi_t \right) dt - c_t dt + W_t \left( \phi_t^m \sigma_m dB_t + \phi_t d\hat{Z}_t \right),
\]
where $\phi_t^m = \psi_t^m + \beta \phi_t$ and $\psi_t = \sigma \phi_t$. Note from above that the expected evolution of the wealth of the individual depends on the filtration parameter $\Lambda$, which represents the amount of information the agent has at any given time. This indicates that $\Lambda$ is expected to play a key role in determining the individual’s portfolio holdings, and it is to this issue we now turn.

From the Hamilton-Jacobi-Bellman equation, we obtain the following:

$$\max_{c, \phi_m, \phi} \alpha^\gamma e^{-\rho t} \frac{c^{1-\gamma}}{1-\gamma} + \frac{\partial}{\partial t} J + W \left( r + \mu \phi_m + \phi \frac{\Lambda}{T - t + \sigma^2} \right) - c \right) J_W$$

$$+ \frac{1}{2} (\phi^2_\sigma^2 \sigma_m^2 + \phi^2) W^2 J_W W - \frac{\Lambda}{T - t + \sigma^2} J_\Lambda + \frac{1}{2} J_{\Lambda \Lambda} - W \partial J_{WA} = 0,$$

with the terminal condition

$$J(t, W_T, \Lambda) = e^{-\rho T} \frac{W_T^{1-\gamma}}{1-\gamma}.$$

We solve for the optimal portfolio strategy by conjecturing that the indirect utility function should have the form

$$J(t, W_t, \Lambda_t) = e^{-\rho t} \frac{W_t^{1-\gamma}}{1-\gamma} f^\gamma(t, \Lambda_t).$$

From the Hamilton-Jacobi-Bellman equation, we obtain the following:

$$\max_{c, \phi_m, \phi} \alpha^\gamma e^{-\rho t} \frac{c^{1-\gamma}}{1-\gamma} + \frac{\partial}{\partial t} J + W \left( r + \mu \phi_m + \phi \frac{\Lambda}{T - t + \sigma^2} \right) - c \right) J_W$$

$$+ \frac{1}{2} (\phi^2_\sigma^2 \sigma_m^2 + \phi^2) W^2 J_W W - \frac{\Lambda}{T - t + \sigma^2} J_\Lambda + \frac{1}{2} J_{\Lambda \Lambda} - W \partial J_{WA} = 0,$$

with the terminal condition

$$J(t, W_T, \Lambda) = e^{-\rho T} \frac{W_T^{1-\gamma}}{1-\gamma}.$$

We solve for the optimal portfolio strategy by conjecturing that the indirect utility function should have the form

$$J(t, W_t, \Lambda_t) = e^{-\rho t} \frac{W_t^{1-\gamma}}{1-\gamma} f^\gamma(t, \Lambda_t).$$
The first order condition for consumption $c$ is given by

$$\alpha^\gamma c^{-\gamma} = W^{-\gamma} f^\gamma,$$

so that

$$c = \alpha \frac{W}{f}. \quad (2.18)$$

As can be seen from the above expression, the consumption of the agent is a known proportion of current wealth.

The first order conditions for the portfolio weights are

$$\mu W J_W + \varphi^m \sigma^2_m W^2 J_{WW} = 0; \quad (2.19)$$

$$\Lambda \frac{T - t + \sigma^2_\epsilon}{T - t + \sigma^2_\epsilon} W J_W + \varphi W^2 J_{WW} - W J_{W\Lambda} = 0. \quad (2.20)$$

This gives

$$\varphi^m_t = \frac{\mu}{\gamma \sigma^2_m}; \quad (2.21)$$

$$\varphi_t = \frac{\Lambda}{\gamma(T - t + \sigma^2_\epsilon)} - (\ln f)_\Lambda. \quad (2.22)$$

It can be seen from above that the optimal holding in the stock depends directly on the current $\Lambda$. The bigger is $\Lambda$, the greater is the value of information and the more aggressive is the position taken in the stock.

The HJB equation can be rewritten as

$$\alpha f^{-1} - \rho + \gamma f^{-1} \frac{\partial}{\partial t} f + r(1 - \gamma) - \alpha(1 - \gamma)f^{-1} + \frac{1}{2}(1 - \gamma) \frac{\mu^2}{\gamma \sigma^2_m}$$

$$+ \frac{(1 - \gamma)\gamma}{2} \left( \frac{\Lambda}{\gamma(T - t + \sigma^2_\epsilon)} - (\ln f)_\Lambda \right)^2 - \frac{\Lambda}{T - t + \sigma^2_\epsilon} \gamma f^{-1} f_{\Lambda}$$

$$+ \frac{1}{2}(\gamma f^{-1} f_{\Lambda\Lambda} + \gamma(\gamma - 1)f^{-2} f_{\Lambda}^2) = 0,$$

or

$$\alpha \gamma - \rho f + \gamma \frac{\partial}{\partial t} f + r(1 - \gamma)f + \frac{1}{2}(1 - \gamma) \frac{\mu^2}{\gamma \sigma^2_m} f$$
\begin{align*}
&\alpha \gamma + L f(t, \Lambda; T) = 0; \\
&f(T, \Lambda; T) = 1, (2.23)
\end{align*}

where

\begin{align*}
Lf &= -\rho f + \gamma \frac{\partial}{\partial t} f + r(1 - \gamma) f + \frac{1}{2}(1 - \gamma) \frac{\mu^2}{\gamma \sigma_m^2} f \\
&+ \frac{1 - \gamma}{2\gamma} \left( \frac{\Lambda}{T - t + \sigma_t^2} \right)^2 f - \left( \frac{\Lambda}{T - t + \sigma_t^2} \right) f_{\Lambda} + \frac{1}{2} \gamma f_{\Lambda\Lambda}. \tag{2.25}
\end{align*}

**Proposition 5** Suppose that \( g(t, \Lambda; s, T) \) satisfies

\begin{align*}
Lg(t, \Lambda; s, T) &= 0; \tag{2.26} \\
g(s, \Lambda; s, T) &= 1, \tag{2.27}
\end{align*}

then

\[ f(t, \Lambda; T) = \alpha \int_t^T g(t, \Lambda; s, T) ds + g(t, \Lambda; T, T). \]

Proof. It is obvious that \( \alpha \int_t^T g(t, \Lambda; s, T) ds + g(t, \Lambda; T, T) = g(T, \Lambda; T) = 1 \) so that the terminal condition is satisfied. Furthermore,

\begin{align*}
L \left( \alpha \int_t^T g(t, \Lambda; s, T) ds + g(t, \Lambda; T, T) \right) &= -\alpha \gamma g(t, \Lambda; t, T) + \alpha \int_t^T Lg(t, \Lambda; s, T) ds + Lg(t, \Lambda; T, T) - \alpha \gamma. \tag{2.28}
\end{align*}
where the first term is from $\gamma \frac{\partial}{\partial t}$ on the lower integration limit.

Now we need to solve the following PDE

$$
-\rho g(t, \Lambda; s, T) + \gamma \frac{\partial}{\partial t} g(t, \Lambda; s, T) + r(1 - \gamma) g(t, \Lambda; s, T) \\
+ \frac{1}{2} (1 - \gamma) \frac{\mu^2}{\gamma \sigma_m^2} g(t, \Lambda; s, T) + \frac{1 - \gamma}{2\gamma} \left( \frac{\Lambda}{T - t + \sigma^2} \right)^2 g(t, \Lambda; s, T) \\
- \left( \frac{\Lambda}{T - t + \sigma^2} \right) g_{\Lambda}(t, \Lambda; s, T) + \frac{1}{2} \gamma g_{\Lambda \Lambda}(t, \Lambda; s, T) = 0;
$$

$$
g(s, \Lambda; s, T) = 1. \tag{2.29}
$$

Let $g(t, \Lambda; s, T) = e^{a(t; s, T) + \frac{1}{2} b(t; s, T) \Lambda^2}$. This reduces to the following ODE

$$
-\rho + \gamma \frac{\partial}{\partial t} a + r(1 - \gamma) + \frac{1}{2} (1 - \gamma) \frac{\mu^2}{\gamma \sigma_m^2} + \frac{1}{2} \gamma b = 0;
$$

$$
\gamma \frac{\partial}{\partial t} b + \frac{1 - \gamma}{\gamma} \left( \frac{1}{T - t + \sigma^2} \right)^2 - \frac{2b}{T - t + \sigma^2} + \gamma b^2 = 0;
$$

$$
a(s; s, T) = 0;
$$

$$
b(s; s, T) = 0.
$$

Let $d = (T - t + \sigma^2) \gamma b$ and $\tau = \ln(T - t + \sigma^2)$. We have

$$
-\frac{\partial}{\partial \tau} d + \frac{1 - \gamma}{\gamma} + \left( 1 - \frac{2}{\gamma} \right) d + \frac{d^2}{\gamma} = 0.
$$

The solution is given by

$$
\gamma a(t; s, T) = \left( -\rho + (1 - \gamma) \left( r + \frac{\mu^2}{2\gamma \sigma_m^2} \right) \right) (s - t) + \frac{1}{2} \ln \left( \frac{T - t + \sigma^2}{T - s + \sigma^2} \right) \\
- \frac{1}{2} \gamma \ln \left( \frac{s - t}{\gamma (T - s + \sigma^2)} + 1 \right);
$$

$$
\gamma b(t; s, T) = \frac{1 - \gamma}{(T - t + \sigma^2)} \left[ \frac{s - t}{\gamma (T - s + \sigma^2)} \right].
$$
The function \( f \) is given by

\[
f(t, \Lambda; T) = \alpha \int_t^T e^{a(t; s, T) + \frac{1}{2}b(t; s, T)\Lambda_s^2} ds + e^{a(t; T, T) + \frac{1}{2}b(t; T, T)\Lambda_t^2}.
\] (2.30)

The optimal portfolio weight is given by

\[
\varphi_t^* = \left( \frac{1}{\gamma(T - t + \sigma_t^2)} - \frac{\alpha \int_t^T b(t; s, T)e^{a(t; s, T) + \frac{1}{2}b(t; s, T)\Lambda_s^2} ds + b(t; T, T)e^{a(t; T, T) + \frac{1}{2}b(t; T, T)\Lambda_t^2}}{\alpha \int_t^T e^{a(t; s, T) + \frac{1}{2}b(t; s, T)\Lambda_s^2} ds + e^{a(t; T, T) + \frac{1}{2}b(t; T, T)\Lambda_t^2}} \right) \Lambda_t
\]

\[
= \left( \frac{1}{\gamma(T - t + \sigma_t^2)} - b(t; T, T) + \frac{\alpha \int_t^T (b(t; T, T) - b(t; s, T))e^{a(t; s, T) + \frac{1}{2}b(t; s, T)\Lambda_s^2} ds}{\alpha \int_t^T e^{a(t; s, T) + \frac{1}{2}b(t; s, T)\Lambda_s^2} ds + e^{a(t; T, T) + \frac{1}{2}b(t; T, T)\Lambda_t^2}} \right) \Lambda_t
\]

\[
= \left( \frac{1}{\gamma(T - t + \sigma_t^2)} + \frac{\mu}{\gamma \sigma_m^2} - \beta \varphi_t^* \right) \Lambda_t.
\] (2.31)

\[
\varphi_t^{ms} = \frac{\mu}{\gamma \sigma_m^2} - \beta \varphi_t^*.
\]

For proving Part 2 of Proposition 2, note that without private information,

\[
a \equiv a_0 = \left( -\rho + (1 - \gamma) \left( r + \frac{\mu^2}{2\gamma \sigma_m^2} \right) \right) (s - t)
\]

and \( b = 0 \). Let \( f_0 \) be the value of \( f \) that corresponds to \( a = a_0 \) and \( b = 0 \). Note that when \( \gamma < 1 \), \( f(t, \Lambda; T) > f_0(t; T) \). Therefore, \( \frac{aW}{T} < \frac{aW}{f_0} \). When \( \gamma > 1 \), \( f(t, \Lambda; T) < f_0(t; T) \). Therefore, \( \frac{aW}{T} > \frac{aW}{f_0} \). Thus the informed agent with \( \gamma < 1 \) will consume a greater fraction of his wealth than the agent with \( \gamma > 1 \). This completes the proofs of Propositions 1 through 2.

**Proof of Proposition 3:** First, suppose that \( \gamma < 1 \), then, it can be easily proved that \( a(t) > 0 \) and \( b(t) > 0 \).

\[
\alpha \int_t^T e^{a(t; s) + \frac{1}{2}b(t; s)\Lambda_s^2} ds + e^{a(t; T) + \frac{1}{2}b(t; T)\Lambda_t^2} > \alpha \int_t^T e^{a_0(t; s)} ds + e^{a_0(t; T)}.
\]
Therefore,
\[ f(t, \Lambda; T) > f_0(t; T) \]
and
\[ R = \left( \frac{f(t, \Lambda; T)}{f_0(t; T)} \right)^\frac{\gamma}{\gamma - 1} > 1. \]

Second, consider the case \( \gamma > 1 \). In this case, it can be easily proved that \( a(t) < 0 \) and \( b(t) < 0 \).

\[ \alpha \int_t^T e^{a(t; s) + \frac{1}{2} b(t; s) \Lambda_t^2} ds + e^{a(t; T) + \frac{1}{2} b(t; T) \Lambda_T^2} < \alpha \int_t^T e^{a_0(t; s)} ds + e^{a_0(t; T)}. \]

Therefore,
\[ f(t, \Lambda; T) < f_0(t; T) \]
and again,
\[ R = \left( \frac{f(t, \Lambda; T)}{f_0(t; T)} \right)^\frac{\gamma}{\gamma - 1} > 1. \]

**The Case of Logarithmic Utility:** In the case of \( \gamma = 1 \), we have logarithmic utility, which requires special mathematical treatment. The utility function can be written as

\[
U = \lim_{\gamma \to 1} E_0 \left[ \int_0^T \alpha e^{-\rho t} \frac{X_t^1 - 1}{1 - \gamma} dt + e^{-\rho T} \frac{W_t^1 - 1}{1 - \gamma} \right] \\
= E_0 \left[ \int_0^T \alpha e^{-\rho t} \ln c dt + e^{-\rho T} \ln W_T \right] \\
= \lim_{\gamma \to 1} E_0 \left[ \int_0^T \alpha e^{-\rho t} \frac{X_t^1 - 1}{1 - \gamma} dt + e^{-\rho T} \frac{W_t^1 - 1}{1 - \gamma} \right] - \int_0^T \frac{\alpha e^{-\rho t}}{1 - \gamma} dt - \frac{e^{-\rho T}}{1 - \gamma}. 
\]

So
\[
J = \lim_{\gamma \to 1} \frac{W_t^{1 - \gamma}}{1 - \gamma} \left( \int_t^T \alpha e^{a(t; s) + \frac{1}{2} b(t; s) \Lambda_t^2} ds + e^{a(t; T) + \frac{1}{2} b(t; T) \Lambda_T^2} \right)^\gamma - \int_t^T \frac{\alpha e^{-\rho(s-t)}}{1 - \gamma} ds \\
\frac{e^{-\rho(T-t)}}{1 - \gamma} \\
= g(t) \ln W + h(t, \Lambda). 
\]
Using a Taylor expansion around $\gamma = 1$ and denoting
\[ f_1 = f(\gamma = 1) = \alpha \int_t^T e^{-\rho(s-t)} ds + e^{-\rho(T-t)}. \]

The indirect utility is
\[ J = -\left( f_1 \ln f_1 + \frac{\partial f}{\partial \gamma}(\gamma = 1) \right) + f_1 \ln W + \alpha \ln \alpha \int_t^T e^{-\rho(s-t)} ds. \]

Noting that
\[
\begin{align*}
\lim_{\gamma \to 1} a(t, s; T) &= -\rho(s-t) + (1-\gamma) \left( r + \frac{\mu^2}{2\sigma^2_m} - \rho \right) (s-t) \\
&+ \frac{1}{2} (1-\gamma) \left[ \ln \left( \frac{T-t + \sigma^2}{T-s + \sigma^2} \right) - \frac{s-t}{T-t + \sigma^2} \right] \\
&= -\rho(s-t) + (1-\gamma) A_0 + (1-\gamma) A_1 \\
\lim_{\gamma \to 1} b(t, s; T) &= (1-\gamma) \frac{s-t}{(T-t + \sigma^2)^2} = (1-\gamma) B_1 \\
\frac{\partial f}{\partial \gamma}(\gamma = 1) &= \alpha \int_t^T \left( -A_0(t, s; T) - A_1(t, s; T) - \frac{1}{2} B_1(t, s; T) \Lambda^2 \right) e^{-\rho(s-t)} ds \\
&+ \left( -A_0(t, T; T) - A_1(t, T; T) - \frac{1}{2} B_1(t, T; T) \Lambda^2 \right) e^{-\rho(T-t)},
\end{align*}
\]
results in the indirect utility being
\[ J = f_1 \ln W - f_1 \ln f_1 + \alpha \int_t^T \left( A_0(t, s; T) + A_1(t, s; T) + \frac{1}{2} B_1(t, s; T) \Lambda^2 \right) e^{-\rho(s-t)} ds + \\
+ \left( A_0(t, T; T) + A_1(t, T; T) + \frac{1}{2} B_1(t, T; T) \Lambda^2 \right) e^{-\rho(T-t)} + \alpha \ln \alpha \int_t^T e^{-\rho(s-t)} ds. \]

A similar result can be obtained by directly solving the HJB equation for the log utility case under the conjecture that $h(t, \Lambda) = a(t) + b(t) \Lambda^2$.

The indirect utility for an informed investor is
\[ J_0 = f_1 \ln W - f_1 \ln f_1 + \alpha \int_t^T A_0(t, s; T) e^{-\rho(s-t)} ds + A_0(t, T; T) e^{-\rho(T-t)} + \]
86
\[ +\alpha \ln \alpha \int_t^T e^{-\rho(s-t)} ds. \]

The (transformed) value of information in this case will be

\[
\ln(R) = \frac{\alpha \int_t^T \left( A_1(t, s; T) + \frac{1}{2} B_1(t, s; T) \Lambda^2 \right) e^{-\rho(s-t)} ds}{f_1} + \frac{\left( A_1(t, T; T) + \frac{1}{2} B_1(t, T; T) \Lambda^2 \right) e^{-\rho(T-t)}}{f_1},
\]

and is always positive.

**Proof of Proposition 4:** The ratio of the utilities from being informed and uninformed is given by

\[
R(t)^{1-\gamma} = \frac{(T-t+1)^{\frac{1}{2}}}{(\frac{T-t+\gamma \sigma^2}{\gamma \sigma^2} + 1)^{\frac{1}{2}} + \frac{1}{2} B_1(t, T; T) \Lambda^2} \exp \left( \frac{(1-\gamma)(T-t) \Lambda^2}{2(T-t+\gamma \sigma^2)(T-t+\gamma \sigma^2)} \right).
\]

Noting that at time \( t \), \( \Lambda_t \) has a mean of zero and a variance of \( T-t+\sigma^2 \),

\[
E \left[ R(t)^{1-\gamma} \right] = \int_{-\infty}^{\infty} \frac{\left( \frac{T-t+\sigma^2}{\gamma \sigma^2} \right)^{\frac{1}{2}}}{\frac{1}{2} \sqrt{2\pi} (T-t+\sigma^2)} \exp \left( \frac{(1-\gamma)(T-t) \Lambda^2}{2(T-t+\gamma \sigma^2)(T-t+\gamma \sigma^2)} - \frac{\Lambda^2}{2(T-t+\gamma \sigma^2)} \right) d\Lambda_t
\]

\[
= \sqrt{2\pi \sigma_t} (T-t+\gamma \sigma^2)^{\frac{\gamma}{2}} \int_{-\infty}^{\infty} \exp \left( -\frac{\gamma}{2} \frac{\Lambda^2}{(T-t+\gamma \sigma^2)} \right) d\Lambda_t
\]

\[
= \left( \frac{T-t}{\gamma \sigma^2} + 1 \right)^{\frac{1-\gamma}{2}}
\]

Equation (2.17) follows directly from the above.
Figure 2.1: Consumption to wealth ratio of informed over that of the uninformed as a function of the signal spread, $\Lambda$.

We assume the propensity to consume, $\alpha$, is 1, the volatility of the signal noise, $\sigma_{\epsilon}$, is 0.3, market risk premium, $\mu$, is 6%, market volatility, $\sigma_m$, is 15%, risk free rate is 4%, the time to the event, $T$, is 1, the idiosyncratic volatility of the stock, $\sigma_s$, is 40%, the subjective discount rate, $\rho$, is 0.2 and $\beta$ is 1.
Figure 2.2: Consumption to wealth ratio of informed over that of the uninformed as a function of risk aversion, $\gamma$.

We assume the propensity to consume, $\alpha$, is 1, the signal spread, $\Lambda$, is 0.5, the volatility of the signal noise, $\sigma_\epsilon$, is 0.3, market risk premium, $\mu$, is 6%, market volatility, $\sigma_m$, is 15%, risk free rate is 4%, the time to the event, $T$, is 1, the idiosyncratic volatility of the stock, $\sigma_s$, is 40%, the subjective discount rate, $\rho$, is 0.2 and $\beta$ is 0.6.
We assume the propensity to consume, $\alpha$, is 1, the signal spread, $\Lambda$, is 0.5, market risk premium, $\mu$, is 6%, market volatility, $\sigma_m$, is 15%, risk free rate is 4%, the time to the event, $T$, is 1, the idiosyncratic volatility of the stock, $\sigma_s$, is 40%, the subjective discount rate, $\rho$, is 0.2 and $\beta$ is 1.
Figure 2.4: Holding in stock as a function of the propensity to consume, $\alpha$.

We assume the signal spread, $\Lambda$, is 0.5, market risk premium, $\mu$, is 6%, market volatility, $\sigma_m$, is 15%, risk free rate is 4%, the time to the event, $T$, is 1, the idiosyncratic volatility of the stock, $\sigma_s$, is 40%, the subjective discount rate, $\rho$, is 0.2 and $\beta$ is 1.
Figure 2.5: Holding in stock and in the market as a function of signal noise variance, $\sigma^2_\epsilon$.

We assume the propensity to consume, $\alpha$, is 1, the signal spread, $\Lambda$, is 0.5, market risk premium, $\mu$, is 6%, market volatility, $\sigma_m$, is 15%, risk free rate is 4%, the time to the event, $T$, is 1, the idiosyncratic volatility of the stock, $\sigma_s$, is 40%, the subjective discount rate, $\rho$, is 0.2 and $\beta$ is 1.
Figure 2.6: Ex-post value of information as a function of $T$.

We assume the propensity to consume, $\alpha$, is 1, the signal spread, $\Lambda$, is 0.5, market risk premium, $\mu$, is 6%, market volatility, $\sigma_m$, is 15%, risk free rate is 4%, the time to the event, $T$, is 1, the idiosyncratic volatility of the stock, $\sigma_s$, is 40%, the subjective discount rate, $\rho$, is 0.2 and $\beta$ is 0.6.
Figure 2.7: Ex-ante value of information as a function of propensity to consume, $\alpha$.

We assume the signal spread $\Lambda$, is 0.5, market risk premium, $\mu$, is 6%, market volatility, $\sigma_m$, is 15%, risk free rate is 4%, the time to the event, $T$, is 1, the idiosyncratic volatility of the stock, $\sigma_s$, is 40%, the subjective discount rate, $\rho$, is 0.2 and $\beta$ is 1.
Figure 2.8: Initial holdings in stock, in response to private information on earnings surprise, as a function of signal noise standard deviation, $\sigma_\epsilon$.

Calibrated to Campbell, Lo and MacKinlay (1997): the time to the event is 21 days, $T = 21/250$. For a positive surprise, $L$ is calibrated as $0.01966 * 250 / (21 * 0.25) = 0.09362$, and the corresponding number for a negative surprise is $-0.7329$. We assume the propensity to consume, $\alpha$, is 0.2, market risk premium, $\mu$, is 8%, market volatility, $\sigma_m$, is 12%, risk free rate is 4%, the idiosyncratic volatility of the stock, $\sigma_s$, is 25%, the subjective discount rate, $\rho$, is 0.1 and $\beta$ is 0.8. Risk aversion, $\gamma$, is 3.
Figure 2.9: Ex-ante value of information as a function of investment horizon, $T'$. We assume the propensity to consume, $\alpha$, is 1, the signal spread $\Lambda$, is 0.5, market risk premium, $\mu$, is 6%, market volatility, $\sigma_m$, is 15%, risk free rate is 4%, the time to the event, $T$, is 1, the idiosyncratic volatility of the stock, $\sigma_s$, is 40%, the subjective discount rate, $\rho$, is 0.2 and $\beta$ is 1.
Figure 2.10: Holding in stock as a function of investment horizon, $T'$. 

We assume the propensity to consume, $\alpha$, is 1, the signal spread $\Lambda$, is 0.5, market risk premium, $\mu$, is 6%, market volatility, $\sigma_m$, is 15%, risk free rate is 4%, the time to the event, $T$, is 1, the idiosyncratic volatility of the stock, $\sigma_s$, is 40%, the subjective discount rate, $\rho$, is 0.2 and $\beta$ is 1.
Table 2.1: Value of Insider Information About Earnings Surprises

This table reports the value of private information in the case of insiders anticipating a firm’s earnings surprise. The value of private information is defined in the text as the ratio of the certainty equivalent with the information signal to that without the signal. Campbell, Lo and MacKinlay (1997) show that the 21 day cumulative abnormal return (CAR) in advance of a positive earnings surprise is 1.966%. The analogous number for a negative surprise is -1.539%. Since the CAR can be viewed as the cumulative realization of $\sigma_s dZ_t$, the signal $L$ can be calibrated as the cumulative abnormal return scaled by $\sigma_s$, the idiosyncratic volatility of the stock. We choose $\sigma_s = 0.25$ in line with Campbell et al. (2001). For a positive surprise, $L$ is calibrated as $0.01966 \times 250 / (21 \times 0.25) = 0.09362$, and the corresponding number for a negative surprise is -0.7329.

<table>
<thead>
<tr>
<th>Signal Noise ($\sigma_s$)</th>
<th>Good News</th>
<th>Bad News</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>27.51</td>
<td>9.81</td>
</tr>
<tr>
<td>0.20</td>
<td>4.31</td>
<td>2.57</td>
</tr>
<tr>
<td>0.30</td>
<td>1.83</td>
<td>1.47</td>
</tr>
<tr>
<td>0.40</td>
<td>1.31</td>
<td>1.19</td>
</tr>
<tr>
<td>0.50</td>
<td>1.14</td>
<td>1.09</td>
</tr>
<tr>
<td>0.60</td>
<td>1.07</td>
<td>1.05</td>
</tr>
<tr>
<td>0.70</td>
<td>1.04</td>
<td>1.03</td>
</tr>
<tr>
<td>0.80</td>
<td>1.03</td>
<td>1.02</td>
</tr>
<tr>
<td>0.90</td>
<td>1.02</td>
<td>1.01</td>
</tr>
<tr>
<td>1.00</td>
<td>1.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>
CHAPTER 3

Choice Architecture and Retirement Saving Plans

In this chapter, we apply basic principles from the domain of design and architecture to choices made by employees saving for retirement. Three of the basic principles of design we apply are: (1) there is no neutral design, (2) design does matter, and (3) many of the seemingly minor design elements could matter as well. Applying these principles to the domain of retirement savings, we show that the design of retirement saving vehicles has a large effect on saving rates and investment elections, and that some of the minor details involved in the architecture of retirement plans could have dramatic effects on savings behavior. We conclude this chapter by discussing how lessons learned from the design of objects could be applied to help people make better decisions, which we refer to as “choice architecture.”
3.1 Introduction

On March 28, 1979, the Unit 2 nuclear power plant on the Three Mile Island Nuclear Generating Station in Dauphin County, Pennsylvania, suffered a core meltdown. In the investigation that followed, it became clear that a valve that was supposed to regulate the flow of cooling water had failed. The operators sent a control signal to remotely shut the valve, and when they received an indication that the signal had been sent, they assumed that the valve was indeed shut. An actual “positive feedback” lamp indicating the true position of the valve did not exist, so the operator had no way of verifying whether the signal was received and the necessary actions taken. Such lamp was deemed expendable during the construction of the facility to save time and money. As a result of this design error, operators were unaware that the valve was not turned off, that the cooling water continued to pour out, and that the reactor’s core continued to overheat and eventually melted down. Even though initial reports blamed “human error,” subsequent investigations found the design of controls equally at fault. They determined that ringing alarms and flashing warning lights left operators overwhelmed by information, much of it irrelevant, misleading or incorrect. ¹

The Three Mile Island incident is one example of how a faulty design can lead even highly qualified decision makers to devastating results. Although managing a retirement portfolio is not a national mission-critical operation, a financial meltdown can be just as painful to an individual as a plant meltdown is to the masses. In this article, we propose that the design of nuclear plant control rooms, everyday objects and retirement saving vehicles share similar properties.

There are two crucial factors to consider. First, everything matters. Tiny

¹A concise description of the event can be found in the US Nuclear Regulatory Fact Sheet at http://www.nrc.gov/reading-rm/doc-collections/fact-sheets/3mile-isle.html
details, from the color of an alert lamp to the size of the font can influence choices. Second, since everything matters, it is important for those who design choice environments, who Thaler and Sunstein (2008) call “choice architects” to take human factors into account. Choice architecture is particularly important in domains such as retirement savings where most of the decision makers are unsophisticated.

Prior research in the domain of retirement savings has illustrated the potential role of improved choice architecture. Madrian and Shea (2001), for example, showed that the choice of default has a dramatic effect on savings behavior. They have studied several plans that changed the default so that employees who take no action are automatically enrolled into the retirement savings plan. It is important to note, however, that freedom of choice is preserved as employees could always opt-out of the retirement plan and are not in any way forced to save. In one of the plans studied, the percentage of employees saving for retirement increased from 49 percent to 86 percent as the default was changed to automatically enrolling employees into the plan.

Other studies have also documented that design does matter. Benartzi and Thaler (2001), for example, showed that the menu of investment funds offered to employees affects their risk-taking behavior. In particular, some employees follow a naive diversification strategy of spreading their money equally across funds, something they have dubbed the 1/n rule. As a result of using the 1/n rule (or a variant of this rule), a plan offering a bond fund, a small cap stock fund and a large cap stock fund might result in employees leaning toward an allocation of two thirds in stocks. In comparison, a plan with a money market fund, a bond fund and a diversified stock fund might result in just one third
Iyengar and Kamenica (2006) documented that the size of the menu of funds also affects savings behavior. They studied a cross-section of retirement savings plans, some offering as few as two funds and others with as many as 59 funds. They estimate that the addition of 10 funds to the menu of choices decreases participation in the plan by two percent, as some employees might be overwhelmed by the degree of choice.

The intuitive principle that many minor design elements could end up being important also applies to retirement saving vehicles. Benartzi and Thaler (forthcoming), for example, show that the number of lines displayed on the investment election form could have the unintended consequence of influencing the number of funds people choose. In one experiment they conducted, visitors to the Morningstar.com website (an online provider of financial information) were presented with an investment election form that had either four or eight lines displayed. Note that those who were presented with four lines could still select more than four funds by simply clicking on a link to the form with eight lines. Benartzi and Thaler found that only 10 percent of those presented with four lines ended up picking more than four funds versus 40 percent for those who saw eight lines on their form to begin with. In other words, the graphic designer who creates the investment election form could accidentally influence the number of funds people pick.

In this chapter, we provide new evidence on choice architecture in the domain of retirement savings plans. We focus on two timely design issues related to the

\footnote{Huberman and Jiang (2006) extended the analysis in Benartzi and Thaler and showed that employees are more likely to use the $\frac{1}{n}$ heuristic when the number of funds is small and when 100 percent is divisible by $n$. For example, only five percent of those selecting nine funds use an approximately equal allocation across the nine funds, whereas 53 percent of those using 10 funds use an equal allocation.}
Consistent with the work of Madrian and Shea (2001), we find that inertia plays a crucial role in choice architecture. In particular, we find that when the escalator program is set as an opt-in program, about 15 - 25 percent of new hires sign up for the program. In contrast, when employees are automatically enrolled in the escalator program, only 16.5 percent opt out and the remaining 83.5 percent end up in the escalator program. We also find that seemingly minor design elements do matter in the context of escalator programs. For example, we document that employees prefer to pre-commit to save more next January as opposed to say next February or next March. In the spirit of New Year’s resolutions, people seem to think that January is a good time to start exerting willpower.

The second design issue we explore has to do with portfolio solutions. In recent years, fund providers have come up with one-stop portfolio solutions to assist employees with the complicated task of fund selection. One solution offered
by fund providers is risk-based funds. These funds are often labeled conservative, moderate or aggressive, and employees are expected to pick the one fund that matches their risk preferences. A distinctive feature of risk-based funds is that they keep a constant asset allocation and do not reduce their equity exposure as people get older. A competing solution offered by fund providers is retirement date funds. These funds are often labeled 2010, 2020, 2030 and 2040, where the labels correspond to the expected retirement date. Unlike risk-based funds, retirement date funds decrease their equity exposure as people approach retirement. In the case of retirement date funds, employees who are looking for a simple portfolio solution should pick the fund that matches their expected retirement date.

One might view the packaging of bond funds and stock funds into one-stop portfolio solutions as inconsequential, since individuals still have access to the underlying bond funds and stock funds to select the mix of funds they truly prefer. However, we find that one-stop portfolio solutions increase equity market participation by about three percentage points. More importantly, the effect is larger for lower-income employees, hence it reduces the well-documented gap in equity market participation between lower-income and higher-income individuals. We also find that retirement date funds strengthen the negative correlation between age and risk-taking behavior. It is important to note that the stronger negative correlation between age and risk taking is observed not only for investors in retirement date funds, but also for the entire population of participants in plans offering retirement date funds. Understanding how the architecture of one-stop portfolio solutions affects investor behavior is essential in light of PPA and the related guidelines by the Department of Labor blessing a spectrum of one-stop portfolio solutions.
The rest of the chapter is organized as follows. In section two, we discuss choice architecture and the effectiveness of escalator programs in increasing saving rates. In section three, we discuss the effects of choice architecture on portfolio choices. In both sections, we provide new evidence that design matters and that seemingly minor design elements could end up being important. We provide concluding remarks in section four.

3.2 Choice architecture and escalator programs

3.2.1 Background information

The worldwide trend toward defined contribution retirement plans has shifted the responsibility for retirement planning from the employer to employees. In most defined contribution plans, employees have to figure out how much to save for retirement and how to invest their funds. Given the difficulty of calculating the “optimal” saving rate as well as self-control problems, it should not come as a surprise that most people are not saving enough to maintain comfortable lifestyle at retirement (Skinner, forthcoming). And as we noted earlier, 68 percent of plan participants agree that their saving rate is “too low” (Choi et al, 2002).

Being interested in helping people reach their stated goal of saving more, we have used the basic psychological principles of hyperbolic discounting, inertia and nominal loss aversion to design a program that helps employees increase their saving rates. The program offers individuals the opportunity to pre-commit to automatic saving increases, which could take place every time someone receives a pay raise, or alternatively, on a set date like every January 1. Of course, participants in the program can always change their mind and either stop the automatic saving increases or quit saving altogether. We have dubbed the program “Save
More Tomorrow™ (hereafter, “SMarT”).

Features of SMarT have been incorporated in PPA, which encourages employers to automatically enroll new and existing employees into their retirement savings plans. The Act prescribes an initial saving rate of at least three percent of pay, an annual increase increment of at least one percent, and a target rate of at least six percent, but no more than ten percent. Employers who follow the above guidelines and provide a generous matching contribution are exempt from the non-discrimination tests (i.e., they do not have to prove that lower-paid employees are benefiting fairly from the retirement plan in comparison to higher-paid employees). Note that the Act allows for saving increases to take place on any date and does not require that saving increases and pay raises be synchronized. Similar legislative initiatives are taking place in the U.K. and New Zealand.

Retirement plan providers have also expressed great interest in automatic saving increases. Vanguard has already made the program available to more than one million employees, Fidelity (2006) reports that 6,000 of its employer clients already offer the program to their employees, and T. Rowe Price and TIAA-CREF, among other providers, have also rolled out similar programs. A survey by Hewitt Associates (2007) indicates that thirty-one percent of plan sponsors already offered the program to their employees in 2006, and that 42 percent of those who did not offer it were likely to offer it in 2007. Similar programs are also being introduced in the U.K. and Australia. The rapid penetration of the program into the marketplace reflects the importance of choice architecture.

The accumulating data on the program suggests dramatic cross-sectional dif-

---

5 “Save More Tomorrow” is a registered trademark of Shlomo Benartzi and Richard H. Thaler. The program is also referred to as the “SMarT” program, “auto-increase” and “contribution escalation.” Firms are more than welcome to use the program free of charge as long as they are willing to share data for research purposes.

6 For a summary of the legislative changes around the world, see Retirement Security Project (2006) and Iwry (2006).
ferences in employee take-up rates. In our original case study with one-on-one financial counseling, take-up rates reached 80 percent (Thaler and Benartzi, 2004). With automatic enrollment into the program, participation rates also reach 80 percent. On the other end of the spectrum, some retirement service providers recently reported take-up rates as low as a few percentage points.

In this section, we attempt to identify the design elements of the program that are most effective at helping people better reach their retirement savings goals. Our research is driven by both theoretical and practical interests. From a theoretical perspective, we are interested in better understanding the psychology of saving. From a practical perspective, we are interested in fine-tuning the program to help more people save more.

We will next describe the psychological principles underlying the program in more detail - i.e., hyperbolic discounting, inertia and nominal loss aversion. As we describe each psychological principle and design element, we also investigate its role in the success of the program. We will also compare each design element to the specific plan design features prescribed by PPA.

3.2.2 Hyperbolic discounting

The first psychological principle that guided us in the design of the program was hyperbolic discounting, which refers to a discount function that “over-values the more proximate satisfaction relative to the more distant ones” (Strotz, 1955, p. 177). Read and Leeuwen (1998), for example, asked subjects to choose between healthy snacks (bananas) and unhealthy snacks (chocolate). When asked one week in advance, only 26 percent of subjects indicated they would choose the unhealthy snack. However, when asked immediately prior to consuming the snack, 70 percent chose the unhealthy snack. This form of present-biased preferences
is characterized by the discount rate increasing as consumption gets closer (see Thaler, 1981, and Loewenstein and Elster, 1992, and Frederick et al, 2002, for additional evidence).

Hyperbolic discounting and present-biased preferences could explain why many of us engage in suboptimal behavior such as excessive eating, lack of exercise and excessive spending. Yet, at the same time, many of us envision we would eat less, exercise more and save more in the not-too-distant future. Hyperbolic agents believe (often wrongly) that doing the right things would be easier in the future, as the temptation to say eat too much would be moderated (see work by Laibson, 1997, and O'Donoghue and Rabin, 1999, 2001, modeling such behavior). To help hyperbolic agents save more, our program invites employees to sign up to save more in the future. However, we do not know the role of this specific feature in the success of the program, nor do we know what should be the time lag between signing up for the program and the effective date of the first saving increase to encourage maximum participation in the program.

The first SMarT case study (Thaler and Benartzi, 2004) offers some insight into the role of hyperbolic discounting in the success of the program. In that case study, they found that 78 percent of those who declined to increase their saving rates right away agreed to do so every time they get a pay raise. This pattern of behavior is consistent with hyperbolic discount functions. However, this evidence is more of a joint test of hyperbolic discounting and nominal loss-aversion, as the distant saving increases were synchronized with pay raises and employees never saw their take-home pay decrease. In this work, we provide more direct evidence on the role of hyperbolic discounting.

We explored the role of hyperbolic discounting in several ways. First, we obtained data from Vanguard, a large provider of retirement plan services that
rolled out an automatic increase service called “OneStep Save$^{TM}$” at the beginning of 2004.\footnote{OneStep Save is a registered trademark of Vanguard.} We looked at 65,452 plan participants in 273 plans who were hired when these plans already offered the opportunity to join the OneStep program. Joining the program had to be done via the web or the phone at the employee’s initiative. The results show 15.1 percent of the new employees joined the program. Participation rates vary by plan with an inter-quartile range between 4 and 19 percent.

What makes the Vanguard data of particular interest for our analysis is the fact that almost all individuals had to select the timing of saving increases on their own.\footnote{One could potentially argue that requiring individuals to choose the month of saving increase is inconsistent with the spirit of the program, which is to make saving decisions as simple as possible. Choi et al (2005) provide evidence that simplifying the enrollment process, so that individuals joining a 401(k) plan should only check the “yes” box to a pre-determined combination of saving rate and investment elections, increases participation rates.} While the saving increases take place once a year, the participant still has to select the specific month for the increase to apply as well as the saving increment. Hyperbolic agents are predicted to prefer to increase their saving rate sometime in the future, though theory does not tell us how much of a delay between joining the program and increasing savings individuals would like to have.

Figure 3.1 displays the number of months that have passed between participants signing up for the program and their desired date of saving increase. The figure is based on 49,433 participants who joined the program as of year end 2005, and it reveals some interesting differences across individuals. Some prefer to implement the program sooner rather than later. Specifically, 8.9 percent of participants prefer the saving increase to be implemented within the same month they sign up. However, the remaining 91.1 of participants prefer to postpone saving increases, consistent with hyperbolic discounting. At the extreme, 15.7
percent prefer that the first increase take place exactly one year after signing up, and 10.0 percent of participants would like to wait longer than a year.

We suspect there could also be a time of the year effect. January might be a good candidate, since hyperbolic agents might consider doing the right things “next year.” Figure 3.2 describes the month of increase selected by the program participants. Almost 40 percent of participants actually selected January as the month of increase, and no other month seems to have such a dominating effect. The distribution across months is statistically different from a uniform distribution at the 0.01 level.

While the Vanguard data is consistent with hyperbolic discounting, it does not measure the strength of preferences. For example, would those who postponed their saving increase by one year still join a program that is set to increase saving much earlier? To answer this question, we used data from T. Rowe Price (hereafter TRP), another large provider of retirement plan services. TRP conducted an online survey of plan participants at a large firm during September of 2005. Participants were given a short paragraph describing the program and then asked whether or not they would be interested in signing up for the program. The saving increases were set to take place in “X” months, where X was varied from implementing the increase immediately to postponing implementation by 12 months. Each participant only responded to one of the conditions.

Figure 3.3 displays the intended sign-up rates for the different conditions. Generally, about 30 percent of the participants intend to sign up. However, there is something special about postponing saving increases by 12 months, where the sign up rate is 41 percent ($p < 0.05$). This result is consistent with the Vanguard data, in which delaying the increase by exactly one year was more popular than any other choice.
The results are consistent with a combination of hyperbolic discounting and some type of mental accounting.\(^9\) Models of hyperbolic agents could explain why many participants prefer to postpone saving increases, but it is not obvious why postponing the increases by three months, six months and nine months is equally attractive, yet postponing by 12 months is more attractive. Similarly, it is not clear why postponing to January is more attractive than any other month. We speculate it might have something to do the tradition of turning over a new leaf at the start of the year.

The PPA provides flexibility with respect to the timing of saving increases. Hence, employers could, for example, pick January as the month of implementing saving increases to encourage employee participation. More generally, minor design elements such as the month of the saving increases could end up influencing employee saving behavior.

### 3.2.3 Inertia

The second psychological principle that guided us in the design of the program was inertia, or what Samuelson and Zeckhauser (1988) dubbed the “status quo bias.” Inertia is known to have a dramatic effect on participants’ behavior in defined contribution plans. Typically, inertia prevents individuals from taking the right actions. For example, many participants do not rebalance their portfolios at all (Samuelson and Zeckhauser, 1988) whereas others do not even join the retirement plan, even when it is a virtual arbitrage opportunity (Choi et al, 2004). On the other hand, inertia can also be used in a positive way to enhance plan participation. For example, flipping the default so that employees are automatically enrolled in the plan, unless they take an action to opt out, increases

---

participation rates dramatically (Madrian and Shea, 2001).\textsuperscript{10}

The SMarT program attempts to use inertia in a positive way to help people reach their stated goals of saving more. In particular, once an individual signs up for the Save More Tomorrow program, future saving increases take place automatically unless the individual changes his/her mind and opts out. Another plan design option is to automatically enroll employees into the Save More Tomorrow program. So, unless someone opts out, he/she would automatically be in the program and future saving increases will take place automatically as well. In this work, we explore the effect of automatically enrolling people into our program. The powerful evidence on the role of inertia leads us to hypothesize that default choices would have a significant impact on participation rates in the program.

The first implementation of the program on an opt-out basis took place in 2003 by the Safelite Group, a client of Strong Retirement Services. The program was introduced to employees in June 2003 with an effective saving increase date of July 2003, an annual increment of one percent of pay, and no synchronization between pay raises and saving increases. It is important to note that hyperbolic discounting probably would not play a major role in this setting, as saving increases took place relatively soon after enrolling in the program. And, nominal loss-aversion should not at all play a role in this setting, since saving increases took place on a set date regardless of pay raises. Hence, this is a unique opportunity to identify and focus on the role of inertia in the success of the program.

We have summary statistics on 3,640 employees who were already participating in the 401(k) plan as of May 2003, the month prior to the introduction of the automatic increase program. Ninety-three percent of participants took no action,\textsuperscript{10} Also see work by Johnson and Goldstein (2003) on the effect of defaults on organ donations. They find that countries with explicit consent have about 10 to 20 percent of people make their organs available for donation, whereas countries with implicit consent have about 90 percent of people make their organs available (i.e., only 10 percent opt out).
thus they were automatically enrolled in the program. Six percent have actively opted out of the program, and the remaining one percent of participants used this opportunity to increase their saving rate beyond the automatic increase.

Since the Safelite Group implementation in July of 2003, additional implementations on an opt-out basis have taken place. In our Vanguard dataset, we have 13 plans that have introduced an opt-out version of the program, one in July of 2004 and the rest in 2005.\textsuperscript{11} The opt-out programs cover new hires only, and they are typically set with an initial deferral rate of about three percent of pay and an annual increment of one percent of pay. There is substantial variation in the “cap,” with some plans stopping the increases at five percent and others stopping it at 25, and even 50, percent. There is also substantial variation in the default portfolio choices, with some plans selecting a money market fund and others selecting a balanced fund or retirement date funds. Hence, this opt-out version of the program is more of an autopilot 401(k) plan where enrollment, deferral rates and portfolio choices are all automatically selected on behalf of employees with the option to opt-out.

Figure 3.4 displays the percentage of Vanguard plan participants who take part in the contribution escalator before and after the introduction of automatic enrollment, and it is based on 2,222 new employees who were eligible for the contribution escalator when they were hired. In the 12 months prior to the implementation of automatic enrollment, 25.1 percent opted into the contribution escalator. However, in the 12 months following automatic enrollment, 83.5 percent of the savers were participating in the escalator program. The differences are statistically significant at the 0.01 level. The dramatic change in participation

\textsuperscript{11}About 50 Vanguard clients are in the process of implementing the program on an opt-out basis.
illustrates the power of inertia and the important role of choice architecture.\textsuperscript{12}

One caveat is that the opt-out program was generally introduced in 2005 with the first saving increase scheduled for 2006. Hence, we cannot determine from our data, which ends in 2005, how many participants, if any, opted-out right before the increase. Data from the one plan that introduced the program in 2004 and already had the first increase in 2005 suggests an opt-out rate of just nine percent, so it does not look like participants opt-out right before the first increase.

Another potential caveat is that opt-out programs might “trick” employees into a program they don’t really want. Choi et al (2005) provide some insightful evidence on this issue from two sets of experiments: one having to do with automatically enrolling people into a 401(k) plan at a modest saving rate (though without automatic increases) and the other having to do with requiring employees to make an active choice whether it is to join or not to join the plan. They find that active decision making results in participation rates that are similar to those of automatic enrollment, so it does not look like automatic enrollment tricks people into the plan (at least in their context of a modest, yet constant, saving rate). It is also important to note that there is no way of avoiding setting a default, and it is not clear that the current default of having procrastinators keep their low saving rates is a better one.

3.2.4 Nominal loss-aversion

The third psychological principle that guided the design of our program is nominal loss-aversion. Loss-aversion refers to the fact that the pain associated with losses is about twice the pleasure that is associated with similar magnitude gains

\textsuperscript{12}We have attempted to explore demographic differences between those who opted out and those who did not, but unfortunately, we had very little demographic information on the new hires.
(Tversky and Kahneman, 1992). To the extent that individuals view increased savings and the respective reduction in spending as a loss, loss-aversion predicts that it could be difficult to help people save more. However, the crucial factor for our program is that people tend to evaluate losses relative to some nominal reference point. For example, in a study of perceptions of fairness (Kahneman, Knetch, and Thaler, 1986), subjects were asked to judge the fairness of pay cuts and pay increases. One group of subjects was told that there was no inflation and was asked whether a seven-percent wage cut was “fair.” A majority, 62 percent, judged the action to be unfair. Another group was told that there was a 12-percent inflation rate and was asked to judge the perceived fairness of a five-percent raise. Here, only 22 percent thought the action was unfair. Of course, in real terms the two conditions are virtually identical, but in nominal terms they are quite different.

To ensure that saving increases are not perceived as losses, our program suggests that pay raises and saving increases be synchronized. For example, the program could be designed so that every time an employee receives a pay raise, he/she takes home half the raise and the remaining half is contributed to the retirement plan. This design feature ensures that the take-home amount does not decrease. It is, unfortunately, easier said than done due to some practical implementation issues. For example, very often information on pay raises is received last minute and there is not enough time to update the contribution rate. Hence, we are interested in understanding the role of nominal loss-aversion from both a theoretical perspective but also from a practical perspective. To the extent that nominal loss-aversion does not play an important role in the success of the program, plan sponsors and plan providers could offer employees a simplified program where saving increases take place on a set date regardless of pay raises.

\[13\] See also work by Shafir, Diamond, and Tversky (1997).
Ultimately, we would like to run a field experiment where employees are randomly assigned to one of two conditions. In one condition, employees would be offered the original version of the program where saving increases and pay raises are synchronized, whereas in the other condition saving increases would take place on a set date regardless of pay raises. There are several practical obstacles that make it very difficult to run such a field experiment. First, employers are often reluctant to offer different retirement plan features to different employees due to legal concerns. Second, it is tricky to synchronize saving increases and pay raises, because employees would like to get an advance notice of the forthcoming saving increase, but information on pay raises is often provided last minute.

Given the above-mentioned difficulties of conducting a randomized field experiment, we decided to conduct a survey with the help of Warren Cormier of the Boston Research Group (Cormier, 2006). The survey group included 5,246 retirement plan participants served by half a dozen different vendors. The subjects were interviewed by phone and asked for their interest in joining an automatic saving increase program. One group of subjects were told that saving increases would take place every January and there was no mention of pay raises. Specifically, they were told that:

“Some 401(k) plans offer a new program to make it easy for employees to save more. If you join the program, each January the percentage of your pay that you’re contributing to your plan will automatically increase by 1%, until you reach a savings rate of 15%. So if you are currently contributing 5%, the program would increase your contribution to 6%. Of course, you are in control and can stop the increases at any time.”

Another group of subjects were further told that the saving increases could be synchronized with pay raises. Specifically, they were told that:
“You could also choose to have the amount you’re contributing automatically increase by 1% every time you get a pay raise instead of every January. With this feature, your savings will never cause your take-home pay to go down.”

The results of the survey are displayed in Figure 3.5. Thirty-two percent of the subjects say that they are either very interested or extremely interested in the non-synchronized program that automatically increases their saving rates every January regardless of pay raises. In comparison, 38 percent of the subjects say they are either very interested or extremely interested in the synchronized version of the program that allows for the increases to take place every time a pay raise is received. The difference in the degree of interest in the program is statistically significant at the 0.01 level.

To summarize our findings so far, it seems as though inertia plays the most dominant role in the program, where defaulting employees into the program results in nearly universal participation. This result might not be as trivial and expected as it might seem. While defaults are very powerful, employees do not always stick to the default. Anecdotal evidence from the U.S. indicates that a lot of employees opt out of their defined benefit plans and select the lump sum option. Similarly, Alessandro Previtero shared with us interesting data from Italy, where more than 80 percent of employees opt out of the default investment for their severance package.

Hyperbolic discounting plays a role as well in the success of escalator programs, with a 12-month delay between sign up and saving increases raising projected participation by roughly 10 percent. As to the role of nominal loss-aversion, synchronizing saving increases and pay raises increases the percentage of subjects that are either very- or extremely interested in joining the program by six percent. It does seem, however, that the role of nominal loss-aversion is a second
order effect.

The PPA encourages automatic enrollment with an initial deferral rate of at least three percent of pay. The PPA also encourages automatic increases to a minimum deferral rate of six percent of pay. Employers who follow the prescribed guidelines are exempt from the non-discrimination tests (i.e., they do not have to document that lower-paid employees and higher-paid employees are all benefiting from the plan).

The PPA seems to have incorporated the right design elements. It encourages automatic enrollment and automatic saving increases in line with the research findings on the powerful role of inertia in participants’ behavior. In addition, the PPA provides flexibility on the timing of increases, and it remains silent on the issue of synchronization. The PPA prescribes annual saving increases, but there is no requirement that the increases be synchronized with pay raises. Given the second order effect of nominal loss-aversion in the program and the practical difficulties in synchronizing pay raises and saving increases, mandating synchronization could have been excessively burdensome.\textsuperscript{14}

\section*{3.3 Choice architecture and portfolio choices}

\subsection*{3.3.1 Background information}

Research on participants’ behavior in retirement saving plans indicates that individuals have a hard time saving enough and constructing a well-diversified portfolio (Benartzi and Thaler, forthcoming). Retirement plan providers have

\textsuperscript{14}Other design elements we explored are the annual increase increment and the “cap” (i.e., the rate at which saving increases stop). We found that employees are insensitive to the annual increment being one percent or two percent of pay. Similarly, we found that employees are insensitive to the cap being set at either 10 percent or 20 percent of pay. However, setting an unrealistically high cap tends to demotivate employees and reduce sign-up rates.
attempted to help employees make better portfolio choices by offering simple one-stop solutions. There are at least two types of portfolio solutions in the marketplace, one being risk-based funds (often called lifestyle funds) and the other being retirement date funds (often called lifecycle funds). Risk-based funds maintain a constant level of risk, and they are often labeled “conservative,” “moderate” or “aggressive” to convey their level of risk. Employees who are offered risk-based funds should simply pick the fund that best fits their risk preferences, though we must admit that figuring out your risk preference is easier said than done.

Retirement date funds are different from risk-based funds in that they follow lifecycle investment models rather than a fixed asset allocation. In particular, retirement date funds decrease their risk level as the retirement date approaches. One strategy available to employees who are offered retirement date funds is to simply pick the fund that matches their projected retirement date.

We refer to risk-based funds and retirement date funds as asset allocation funds (though the term we use should not be confused with tactical asset allocation funds that periodically make bets on certain asset classes). Asset allocation funds play an increasingly important role in defined contribution plans. Hewitt Associates (2007) surveyed 146 employers and found that 57 percent offer retirement date funds and 38 percent offer risk-based funds. Policy makers have also expressed interest in asset allocation funds. The Department of Labor has recently issued proposed guidelines on appropriate investments for defined contribution plans in the context of employees who are automatically enrolled into a retirement plan and are defaulted into a an investment or portfolio set by their plan sponsor. The guidelines encourage the use of asset allocation funds.

Given the increasing role of asset allocation funds in retirement plans, we are
interested in exploring the effect of choice architecture in this domain. In particular, we are interested to learn how the packaging of cash, bond funds and stock funds into these one-stop portfolio solutions affects behavior. Since employees can still select any mix of cash, bonds and stocks by using the underlying investment funds to self-construct their own portfolio, one might view such repackaging as inconsequential. However, our data suggests that repackaging and choice architecture do matter. We begin by describing our data and some descriptive statistics on the usage of asset allocation funds. We then analyze the effect of asset allocation funds on both equity market participation and the lifecycle pattern of investing.

3.3.2 Data and descriptive statistics

Our dataset includes about 1.5 million participants in 1,830 defined contribution plans served by Vanguard. The data provides a snapshot of investment elections made by the participants as of December 2005. In particular, we know the total contributions made during December 2005, the amount invested in each of the asset allocation funds, and the percentage of contributions allocated to equities, bonds and cash. The data also includes the following information for most participants: age, gender, plan entry date, account balance, registration for the www.vanguard.com website, as well as proxies for household income and household financial wealth based on the participant’s nine-digit zip code. At the plan level, our data includes indicators for the following plan features: the availability of loans, the inclusion of company stock in the menu of funds, and

---

15 Yamaguchi et al (2007) explore a closely related dataset and find similar results.
16 A company called IXI collects retail and IRA asset data from most of the large financial services companies. IXI aggregates the data from all companies at the nine-digit zip code level and then calculates the average household assets by zip code. On average, there are 10 to 12 households in a nine-digit zip code area. Next, IXI assigns a wealth rank (from 1 to 24) to each area. We narrow the ranks into 5 groups, with the respective ranges displayed in Table 3.1.
access to a brokerage account.

The Vanguard set of risk-based funds includes four LifeStrategy funds, and the set of retirement date funds includes six Target Retirement funds. In most cases, the sets are offered in their entirety in a given plan. The funds are classified by Vanguard under one category called lifecycle funds and are marketed on their website (www.vanguard.com) as “a transparent, simple-to-use solution for identifying and maintaining a proper asset allocation.” Furthermore, according to the website, “The funds are designed as an investment choice for novice investors. They are the ‘one-stop shopping’ choice offering complete diversification in a single fund.”

One of the issues with the data we have is that information on the menu of funds in the plan was recorded as of June 2005, whereas individual investment elections were recorded as of December 2005. To resolve this issue, we decided to determine the type of funds included in the plan by analyzing the contributions made in December 2005. We consider a plan to offer a certain fund if at least one participant made a contribution to that fund. Based on this classification, we find that 520 plans qualify as offering retirement date funds; 811 qualify as offering risk-based funds; and 95 plans offer both types of funds.\(^{17}\)

We follow Benartzi and Thaler (2001, 2002) and examine the allocations of contributions rather than the allocations of accumulated account balances. Whereas finance theory focuses on the allocation of account balances or total wealth, we prefer to study the allocation of flows into the plan. The reason for this choice is that the allocation of account balances is affected by the investment elections participants made many years ago when they joined the plan and subsequent fund performance. As mentioned earlier, few participants rebalance their

\(^{17}\)Unfortunately, the limited data we have about the plans in our sample does not allow us to determine if adoption of such funds are related to certain plan characteristics.
portfolio allocations. Another issue to consider is that retirement date funds are a relatively recent addition to the menu of funds available to plan participants. Hence, we focus on the allocations made by participants who joined in the last two years whom we dub *new participants*. Participants joining in the last two years are more likely to have been offered retirement date funds when they made their “critical” first selection. We also note that retirement date funds have become available in plans mostly during 2004 and 2005, while risk-based funds have been offered for longer. Some of the new participants in plans that offer retirement date funds did not have the possibility of investing in them when they joined. That is not the case for plans that offer risk-based funds.

We begin our analysis by exploring who uses asset allocation funds and how they are being used. We found that 37 percent of plan participants who are offered retirement date funds use them. However, usage of risk-based funds is somewhat higher, with 48 percent of those who are offered risk-based funds using them. We suspect that the lower usage of retirement date funds is related to them being newer.\(^\text{18}\)

We also examined the cross-sectional variation in adoption rates and found that women, younger employees and those with lower monthly contributions, income and wealth are more likely to use retirement date funds. In particular, women are 6.0 percent more likely to use retirement date funds than men are; employees in their 20s are 8.4 percent more likely to use these funds than employees in their 60s; and, employees with the lowest contributions are 6.2 percent more likely to use these funds than those with the highest contributions. Similar

\(^\text{18}\)One concern is that some plans might have offered asset allocation funds as the default investment option, and we know this could have a strong effect on take up rates. We believe our results are unlikely to be affected by this issue for a couple of reasons. First, we would have expected far higher take up rates, had asset allocation funds been used as defaults. Second, most plan sponsors did not use asset allocation funds as defaults prior to the PPA.
patterns emerge in the adoption of risk-based funds.

The descriptive statistics suggest that retirement date funds cater to demographic groups who are less knowledgeable about investing. For example, Dwyer, Gilkeson and List (2002) and Lusardi and Mitchell (2005) document that women are less knowledgeable about financial matters than men are, and Kotlikoff and Bernheim (2001) find positive correlation between income and financial literacy. According to Vanguard’s website, retirement date funds were “designed as an investment choice for novice investors,” and our data suggests that they do serve this purpose.

In terms of the way in which asset allocation funds are being used, we investigated which employees tend to use them exclusively as a one-stop solution. We find the same demographic groups - that is, women and those with lower account balances or monthly contributions - are more likely to use them exclusively. While it is perfectly sensible for novice investors to use asset allocation funds as a one-stop solution, one wonders why the seemingly more sophisticated men and wealthier employees are not using them as the one-stop solution they are designed to be. We checked whether more sophisticated investors have adopted a “core plus” strategy, where they invest most of their funds in an asset allocation fund, but then have a small tilt toward a more targeted investment such as an international fund. We find little of this type of behavior. In particular, just half (53 percent) of the investors in asset allocation funds invest all their contributions in these funds. Of the remaining 47 percent, four out of five investors place less than half of their contributions in asset allocation funds, precluding the notion that asset allocation funds serve as the building block in a “core plus” strategy. We speculate that investors might fear investing in just one fund, not realizing that asset allocation funds are in fact well-diversified blends of several different
3.3.3 Asset allocation funds and equity market participation

In this section, we analyze the effect of offering asset allocation funds to plan participants on their exposure to equity markets. As long as the equity risk premium is positive and there are no transaction costs, theory predicts that investors would own at least a small amount of equities for diversification purposes. However, papers like Mankiw and Zeldes (1991) and Ameriks and Zeldes (2004) show that this is not the case. Many US households have no exposure to equities at all. In particular, Vissing-Jørgensen (2003) reports a large difference in participation rates between low and high net worth households. By her definitions, just 18 percent of the former group participates in the equity markets, while 93 percent of the latter group owns stocks directly or indirectly. In this section, we study the effect of asset allocation funds on equity market participation.

Table 3.2 displays the fraction of plan participants who own equities in their retirement account. We provide the results for plans that (a) do not offer asset allocation funds, (b) offer risk-based funds, and (c) offer retirement date funds. Several patterns emerge from the results. First, we observe a positive correlation between various measures of wealth and equity market participation, which is consistent with earlier studies in this area. Second, asset allocation funds, be it risk-based funds or retirement date funds, increase equity market participation among those with lower income and account balances. Third, asset allocation funds do not affect equity market participation among the wealthiest. Since asset allocation funds increase equity market participation for lower-income individuals only, these funds tend to close the gap in stock ownership between lower- and higher-income participants. Specifically, we find that asset allocation funds cut
the “participation gap” in approximately half. This is true whether we sort
individuals on their contributions, account balances or income. For example, the
participation gap between those with the lowest and highest plan balances is 20.8
for plans not offering asset allocation funds. That gap, however, decreases to 9.4
percent and 8.7 percent for plans offering risk-based funds and retirement date
funds, respectively.

We further run a probit regression to explain equity market participation with
participant and plan attributes. The regression model is given in the following
equation.

\[
Equity_{ij} = \alpha + \beta \times [Contributions_{ij}|HasAA_j] + \epsilon_{ij}
\]

\(Equity_{ij}\) is an indicator for whether or not individual \(i\) in plan \(j\) holds any
equity in his/her portfolio; \(Contributions\) is the log of the participant’s total
contributions in December 2005; and, \(HasAA_j\) is an indicator for whether or
not plan \(j\) offers any type of asset allocation funds. The parameter estimates are
displayed in Table 3.3 with errors clustered at the plan level following Wooldridge
(2003).

The results confirm that participation increases with contributions. Calcula-
tions not reported in the table indicate that doubling the monthly contributions
to the plan increases the likelihood of the participant owning stocks by 5.2 per-
cent. More interestingly, asset allocation funds increase equity market partici-
pation by 3.1 percent. And, the relationship between contributions and equity
market participation is diminished for plans with asset allocation funds, as indi-
cated from the significantly negative coefficient on the interaction term between
asset allocation funds and contribution level. This latter result is consistent with
the univariate analysis showing that asset allocation funds raise participation in
equity markets among lower-income individuals, hence closing the gap in equity market participation between low- and high-contributors to the plan.

Why does the inclusion of asset allocation funds in the plan’s menu affect equity market participation? Moreover, why does it increase participation among lower-paid employees? One reason could be that these funds reduce participation costs in the equity market, either in terms of fees or by reducing the psychic costs of choosing a fund. In the case of Vanguard, there is no difference in fees since the Vanguard retirement date funds charge the same fees as the underlying funds they own. We thus favor the view that the presence of these funds reduces psychic costs.

Other research also supports the psychic costs argument. Charles Schwab, for example, highlights the time-saving argument on their website (www.schwab.com) by asking “Are you looking for a way to reach your retirement goals, but do not have the time to actively manage your portfolio?” Vissing-Jorgensen (2003) uses the psychic costs argument to explain the participation gap between individuals with low and high account balances. In her model, there is a fixed cost of learning about equity investments, measured as \(X\) number of hours. Since wealthier individuals could earn more dollars from participating in the equity market, they can afford the fixed costs of learning about stocks. We agree that wealthier individuals could earn more dollars from participating in equity markets, but it also costs them more to spend \(X\) hours learning about stocks, since they earn a higher hourly wage. As a result, it is not obvious that wealthier individuals have more of an incentive to participate in equity markets than lower-paid individuals. And therefore, it is unlikely that the implicit costs argument drives our results.

Another explanation for equity market non-participation is offered by Barberis, Huang and Thaler (2006). They suggest that “narrow framing,” the ten-
dency to evaluate the components of one’s portfolio rather than the overall port-
folio, could magnify the risk of investing in stocks. In our setting, the narrow
framing hypothesis would imply that some participants are wary of holding eq-
uity funds even when they construct a well-diversified portfolio, because they are
focused on and are averse to experiencing losses in any element of their portfolio.
Asset allocation funds could mitigate narrow framing by making the individual
components of the portfolio less “accessible.” Note, however, that asset allocation
funds could mitigate narrow framing consciously or unconsciously. For example,
investors might be aware that asset allocation funds invest in equities, but find
it palatable since the volatility of stock returns is not segregated. Alternatively,
investors might not even be aware that asset allocation funds invest in equities.

3.3.4 Retirement date funds and lifecycle investment patterns

Despite extensive theoretical work on the relation between investment horizon
and optimal risk-taking behavior, academic “prescriptions” are still mixed on
whether or not there should be a relation between age and portfolio choices as
well as the exact form of the relation. Seminal work by Samuelson (1969) and
Merton (1969) suggests that under certain conditions, the optimal allocation to
the risky asset should remain constant over the lifecycle. In other words, portfolio
choices should be independent of both age and wealth. On the other hand, Bodie,
Merton and Samuelson (1992) and Viceira (2001) incorporate labor income and
human capital as part of one’s overall portfolio and come to a different conclusion.
In particular, they propose that the allocation to the risky asset should decrease
with age. Most financial advisors agree with this advice. One often quoted rule
of thumb is that a person’s asset allocation to equities should be equal to 100
minus his/her age.\textsuperscript{19}

The empirical evidence on actual lifecycle investment patterns is also mixed. Bodie and Crane (1997) find a strong negative relation between age and the fraction of the portfolio invested in stocks. Holden and Derhei (2005) also find a negative relation between equity holdings and age in a large sample of 401(k) plans. Ameriks and Zeldes (2004) use a different research approach of splitting the decision to own any stocks from the decision of how much stock to own. They find that older people are less likely to own stocks, a result driven mainly by plan participants either selling all of their equity holdings at retirement or annuitizing. However, conditional on owning some stock, they find very little correlation between age and the fraction invested in stocks.

Our main interest is whether the inclusion of asset allocation funds in the plan alters the relation between age and risk-taking behavior. Although it is not theoretically clear what the ‘correct’ relation should be, we are able to show that the menu of funds presented to individuals affects employees’ choices. Panel A of Table 3.4 compares the average equity exposure of new participants who held risk-based funds to those who held retirement date funds in December 2005. We break the two samples down by age groups as follows: 20-29, 30-39, 40-49, 50-59, and 60-79. Participants who held both types of funds are excluded.

The relation between age and equity holdings is relatively flat for those investing in risk-based funds. In particular, it goes down from about 69 percent in stocks for those in their 20s and 30s to 62 percent in stocks for those in their 60s. In contrast, investors in retirement date funds exhibit a much stronger correlation between age and risk-taking behavior. In particular, the allocation to stocks decreases from 80 percent for the youngest group to 43 percent for the

\begin{footnote}{Bodie and Crane (1997) describe this rule of thumb and other generally accepted life-cycle investment prescriptions.}

128
oldest group. We test the differences in equity holdings between those who own risk-based versus retirement date funds using both an ANOVA test and a Mann-Whitney Wilcoxon rank test. And, we confirm that investors in retirement date funds hold significantly more equity at the beginning of the lifecycle and significantly less equity at later stages of the lifecycle.

We also compare the two groups to a benchmark group of individuals in plans that offer neither type of funds. We find that investors in risk-based funds display a risk-age relationship that is close to the benchmark while retirement-date investors have on average 11 percent more in stocks in their 20s and 16 percent less in stocks when they are over 60.

The fact that investors who hold retirement date funds display a strong correlation between age and equity holdings is not expected by construction, since investors can utilize other funds to achieve their desired lifecycle pattern of risk taking. However, the more interesting finding is that investors who use risk-based funds do not appear to have a pronounced relation between age and risk-taking behavior. Note that we focus our analysis on new hires, so inertia could not explain the observed pattern.

One caveat is that there could be a selection bias in who chooses asset allocation funds. For example, it is plausible that those who choose retirement date funds prefer to decrease their portfolio risk as they get older, whereas those who choose risk-based funds have a preference for a constant allocation. Furthermore, it is plausible that investors in asset allocation funds would have picked the exact same risk level even if they did not have access to asset allocation funds and had to self-construct their portfolios.

To avoid the selection bias discussed above, we consider the effect of having access to asset allocation funds on all the participants in the plan and not just
those selecting asset allocation funds. Panel B of Table 3.4 displays the results. Again, there is a stronger downward sloping relation between age and risk-taking behavior for plans offering retirement date funds than plans offering risk-based funds or plans offering neither.

In Panel C of Table 3.4, we eliminate participants who do not own any equities from the analysis for two reasons. First, similar to Ameriks and Zeldes (2004), we attempt to separate the decision to have any stock from the choice of how much stock to own. Second, all asset allocation funds in our sample invest in equities, so it seems consistent to compare investors in asset allocation funds to the population of participants investing in stocks. We observe similar patterns for this subsample. Specifically, the average equity exposure in plans offering risk-based funds is 74 percent for the youngest participants and slightly lower at 68 percent for the oldest group of participants. Again, plans offering retirement date funds exhibit a stronger correlation between age and equity exposure, with the youngest participants having 80 percent in stocks versus 61 for the oldest group.

Another caveat is that the menu of funds available to employees could reflect their underlying preferences. For example, one might argue that plan administrators who select retirement date funds had previously identified that participants in their plans are inclined to have portfolios that become more conservative with age. We address this potential bias by looking at the lifecycle pattern of investing for participants in plans offering asset allocation funds who decided to self-construct their portfolios. Specifically, we examine those who hold at least two funds, none of which is an asset allocation fund. Requiring a minimum of two funds increases the likelihood of the participant self-constructing his/her portfolio as opposed to being defaulted into a fund by the employer. We find relatively
flat relation between age and equity exposure for plans offering risk-based funds and retirement date funds (remember, the plans offer asset allocation funds, but our analysis focuses on those not picking the asset allocation funds). This suggests that employees in plans offering risk-based versus retirement date funds are unlikely to be dramatically different a priori.

We further account for the plan-selection bias using a regression model. Table 3.5 reports the results of a censored regression (the lower bound is 0 percent and the upper bound is 100 percent) for the percentage of equity in the portfolio against age, whether the plan has risk-based or retirement date funds, and interaction terms as specified in equation (3.1). Regression results are reported in column (1) without plan level controls and in column (2) with plan level controls. The plan level controls include: portion of female participants, average monthly contribution, average account balance, average tenure, percentage of web-registered users, whether the plan offers loans, company stock or a brokerage account, and the size of the plan using the log number of participants as a proxy. Errors are clustered at the plan level to further account for plan-level effects.

\[
PctEquity_{ij} = \alpha + \beta \times \left[ Age_{ij} \text{ HasRB}_{j}|Age_{ij} \text{ HasRD}_{j}|Age_{ij} \right] + \epsilon_{ij} \tag{3.1}
\]

When the plan does not include asset allocation funds (\(HasRB = HasRD = 0\)), the results indicate a hump-shaped relation between age and equity exposure, with the maximum at about age 37. The coefficients for plans that offer risk-based funds (\(HasRB=1\)) are small and barely significant, indicating that risk-based funds do not alter substantially the relation between age and risk taking. However, retirement date funds change the fitted relationship by making it downward-sloping for ages 25 and above. The slope is also steeper at older ages, as indicated by the negative interaction coefficient.
We derive the marginal effects by calculating the expected change in allocation to equity when decreasing or increasing 10 years of age from the sample average, which is about 38. Compared to the allocation at age 38, the allocation at age 28 is lower by 1.5 percent in plans with no asset allocation funds, is lower by 1.6 percent in plans with risk-based funds, but is higher by 1.8 percent in plans with retirement date funds. Thus, the relationship is downward sloping in early ages only when retirement date funds are offered. On the other hand, the allocation is always downward sloping between ages 38 and 48. The slope is rather flat in the former cases: 2.4 percent and 2.0 percent, and is steeper for plans with retirement date funds: 4.2 percent.

The last caveat we address is that retirement date funds were introduced throughout 2004 and 2005, and some of these were replacing risk-based funds with participants being “mapped” from the risk-based funds into the retirement date funds based on their age. It is plausible that our results are affected by inertia; that is, participants who were mapped to a retirement date fund and never bothered to change their portfolio allocations. To eliminate the possibility that our results may be partially driven by participant inertia, we excluded 52 plans that shifted from risk-based to retirement date funds. The results are virtually identical to those reported earlier. 20

To summarize, choice architecture does affect portfolio choices. The seemingly inconsequential packaging of cash, bonds and stocks into one-stop asset allocation solutions does affect investor behavior. In particular, asset allocation funds enhance equity market participation among lower-paid employees, and as a result, they reduce the equity market participation gap between lower- and higher-paid

---

20We have attempted to use time-series data on the 52 plans that switched from risk-based to retirement date funds. Unfortunately, we did not have a sufficiently large number of new hires in those plans to conduct meaningful analysis.
employees. We also find that the type of asset allocation funds being offered, be it risk-based funds or retirement date funds, affects the lifecycle pattern of investing.

3.4 Conclusion

In this chapter, we highlight the importance of design features of retirement plans. We argue that design does matter and seemingly inconsequential design elements could end up being important. We believe the PPA is an example of good choice architecture. The main design feature of PPA has to do with design for errors or inaction; that is, what happens if people do nothing? In the case of PPA, employees who take no action might still save for retirement as long as their employer follows the PPA prescription of automatically enrolling employees into the plan and escalating their deferral rates periodically.

While PPA has made great use of good choice architecture, it is important to note that there are many domains where choice architecture could be improved. Consider, for example, the Medicare Prescription Drug Program, often referred to as Medicare Part D. There are dozens of different plans offered in each state, making the decision very complicated. The plans actually vary by state, making it impossible for individuals to consult with friends or family members living in other states. There is no spell checker, despite the difficulty of properly spelling the names of some prescription drugs. And, there is no default, unless there are two eligible individuals, in which case they are assigned randomly to one of the plans. Part D is just one of many domains where more research on choice architecture could benefit society.
Figure 3.1: Participants’ Choices of the Delay between Signing-up for the Program and the Effective Date of Saving Increase

The above figure shows the delay between signing up for the escalator program and the effective date of the saving increase. Data was obtained from Vanguard and it consists of 49,433 program participants who joined the program as of year end 2005. For example, 15.7 percent of participants requested that their first saving increase take place 12 months after joining the program.
Figure 3.2: Participants’ Choices of the Month of Saving Increase

The above figure displays the month in which participants have requested their saving rate to increase. Data was obtained from Vanguard and it consists of 49,433 program participants. For example, 37.3 percent of participants requested that their saving rate go up in January.
Figure 3.3: Intended Signup Rates and the Delay between Signing-up for the Program and the Effective Date of Saving Increase

The above figure displays information on an experiment conducted online by T. Rowe Price. Subjects were given a short description of a program offering automatic saving increases and were asked whether or not they would like to join it. The length of time between joining the program and experiencing the saving increase was varied, though each subject was presented with one scenario only. Seven hundred and forty nine subjects participated in the survey.
Figure 3.4: Employee Participation Rates before and after the Implementation of Auto-Enroll Automatic Contribution Escalator

The above figure covers 13 retirement saving plans that introduced automatic enrollment of new employees into the plan as well as into a contribution escalator. The figure displays the fraction of new plan participants who take part in the contribution escalator, and it is based on 2,222 plan participants.
Figure 3.5: The effect of Synchronizing Saving Increases and Pay Raises on the Degree of Interest in Joining the Program

The above figure shows results from an experiment conducted via telephone by the Boston Research Group, where 5,246 retirement plan participants were asked for their level of interest in joining the program. One group of participants was told that saving increases will take place every January (i.e., non-synchronized increases), whereas another group was told that the saving increases could also take place every time they get a pay raise, so that their take-home pay would not go down (i.e., synchronized increases). The chart displays the frequency of responses for the two conditions.
Table 3.1: Usage of Risk-Based and Retirement Date Funds

Sample includes new participants in plans that offer either risk-based or retirement date funds. \( N = 128,540 \) in the former and 74,503 in the latter. *Hold funds* is the percentage of participants that hold the funds, whereas *100% in funds* is the percentage of those who hold only such funds. *Contributions* are for December 2005, and *Wealth* is based on IXI group. ** indicates significance at 1%; * at 10%.

<table>
<thead>
<tr>
<th>Population subgroup</th>
<th>Risk-based funds</th>
<th>Retirement date funds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hold funds (%)</td>
<td>100% in funds</td>
</tr>
<tr>
<td>All</td>
<td>48.2</td>
<td>52.9</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>49.6</td>
<td>53.2</td>
</tr>
<tr>
<td>Male</td>
<td>42.5</td>
<td>39.0</td>
</tr>
<tr>
<td>Difference</td>
<td>-7.1***</td>
<td>-14.2***</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 - 29</td>
<td>50.1</td>
<td>57.7</td>
</tr>
<tr>
<td>30 - 39</td>
<td>48.6</td>
<td>49.7</td>
</tr>
<tr>
<td>40 - 49</td>
<td>47.1</td>
<td>50.3</td>
</tr>
<tr>
<td>50 - 59</td>
<td>46.2</td>
<td>54.1</td>
</tr>
<tr>
<td>60 - 79</td>
<td>43.4</td>
<td>55.8</td>
</tr>
<tr>
<td>Difference (old-young)</td>
<td>-6.7***</td>
<td>-1.9</td>
</tr>
<tr>
<td>Monthly contributions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 - $100</td>
<td>53.9</td>
<td>74.4</td>
</tr>
<tr>
<td>$100 - $200</td>
<td>45.1</td>
<td>54.5</td>
</tr>
<tr>
<td>$200 - $400</td>
<td>47.9</td>
<td>46.1</td>
</tr>
<tr>
<td>$400 - $800</td>
<td>48.2</td>
<td>41.6</td>
</tr>
<tr>
<td>$800+</td>
<td>44.7</td>
<td>39.0</td>
</tr>
<tr>
<td>Difference (high-low)</td>
<td>-9.2***</td>
<td>-35.4***</td>
</tr>
</tbody>
</table>

Table continued on next page ...
<table>
<thead>
<tr>
<th>Population subgroup</th>
<th>Risk-based funds</th>
<th></th>
<th></th>
<th>Retirement date funds</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hold funds (%)</td>
<td>100% in funds</td>
<td>Hold funds (%)</td>
<td>100% in funds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wealth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; $5,000</td>
<td>47.5</td>
<td>56.8</td>
<td>36.5</td>
<td>58.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$5,000 - $25,000</td>
<td>49.8</td>
<td>54.8</td>
<td>37.9</td>
<td>51.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$25,000 - $50,000</td>
<td>49.9</td>
<td>52.6</td>
<td>37.6</td>
<td>48.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$50,000 - $100,000</td>
<td>48.6</td>
<td>50.6</td>
<td>37.4</td>
<td>46.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$100,000+</td>
<td>45.7</td>
<td>48.8</td>
<td>35.4</td>
<td>45.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference (high-low)</td>
<td>-1.8***</td>
<td>-8.0***</td>
<td>-1.1*</td>
<td>-12.6***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table continued from previous page.
Table 3.2: Equity Participation Gap

The table displays the percentage of plan participants that invest in equities. We report equity market participation for participants in plans that (a) offer neither risk-based nor retirement date funds (N = 97,227), (b) offer risk-based funds (N = 119,917), and (c) offer retirement date funds (N = 69,579). Participation gap is the difference in equity market participation.

<table>
<thead>
<tr>
<th>Population subgroup</th>
<th>No asset allocation funds offered</th>
<th>Plan offers risk-based</th>
<th>Plan offers retirement date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly contributions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 - $100</td>
<td>67.8</td>
<td>81.5</td>
<td>75.9</td>
</tr>
<tr>
<td>$100 - $200</td>
<td>77.9</td>
<td>81.1</td>
<td>85.4</td>
</tr>
<tr>
<td>$200 - $400</td>
<td>87.2</td>
<td>89.8</td>
<td>88.7</td>
</tr>
<tr>
<td>$400 - $800</td>
<td>91.9</td>
<td>92.8</td>
<td>92.5</td>
</tr>
<tr>
<td>$800+</td>
<td>93.9</td>
<td>93.9</td>
<td>93.2</td>
</tr>
<tr>
<td>Participation gap (high-low)</td>
<td>26.1</td>
<td>12.4</td>
<td>17.3</td>
</tr>
<tr>
<td>Account balance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 - $1,000</td>
<td>71.7</td>
<td>82.8</td>
<td>82.7</td>
</tr>
<tr>
<td>$1,000 - $2,500</td>
<td>78.2</td>
<td>83.0</td>
<td>83.9</td>
</tr>
<tr>
<td>$2,500 - $5,000</td>
<td>83.9</td>
<td>86.4</td>
<td>87.6</td>
</tr>
<tr>
<td>$5,000 - $10,000</td>
<td>89.1</td>
<td>90.1</td>
<td>88.5</td>
</tr>
<tr>
<td>$10,000+</td>
<td>92.5</td>
<td>92.2</td>
<td>91.4</td>
</tr>
<tr>
<td>Participation gap (high-low)</td>
<td>20.8</td>
<td>9.4</td>
<td>8.7</td>
</tr>
<tr>
<td>Household income</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; $30,000</td>
<td>78.0</td>
<td>83.0</td>
<td>86.0</td>
</tr>
<tr>
<td>$30,000 - $50,000</td>
<td>83.3</td>
<td>84.8</td>
<td>87.7</td>
</tr>
<tr>
<td>$50,000 - $75,000</td>
<td>85.5</td>
<td>87.8</td>
<td>89.9</td>
</tr>
<tr>
<td>$75,000 - $125,000</td>
<td>88.4</td>
<td>89.8</td>
<td>91.4</td>
</tr>
<tr>
<td>$125,000+</td>
<td>90.7</td>
<td>90.6</td>
<td>91.7</td>
</tr>
<tr>
<td>Participation gap (high-low)</td>
<td>12.7</td>
<td>7.6</td>
<td>5.7</td>
</tr>
</tbody>
</table>

Table continued on next page ...

141
<table>
<thead>
<tr>
<th>Population subgroup</th>
<th>No asset allocation funds offered</th>
<th>Plan offers risk-based</th>
<th>Plan offers retirement date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-29</td>
<td>82.9</td>
<td>86.0</td>
<td>87.5</td>
</tr>
<tr>
<td>30-39</td>
<td>85.6</td>
<td>89.0</td>
<td>88.0</td>
</tr>
<tr>
<td>40-49</td>
<td>84.0</td>
<td>87.7</td>
<td>86.5</td>
</tr>
<tr>
<td>50-59</td>
<td>81.5</td>
<td>85.3</td>
<td>84.7</td>
</tr>
<tr>
<td>60-79</td>
<td>76.4</td>
<td>81.7</td>
<td>78.9</td>
</tr>
<tr>
<td>Participation gap (young-old)</td>
<td>6.5</td>
<td>4.3</td>
<td>8.6</td>
</tr>
</tbody>
</table>

Table continued from previous page.
Table 3.3: Equity Market Participation

The table provides regression results for the following probit model:

\[ Equity_{i,j} = \alpha + \beta \cdot [Contributions_{i,j}|HasAA_j] + \epsilon_{i,j} \]

The dependent variable is an indicator equal to 1 if the participant invests in equities. The regressors are the log of the monthly contributions in December 2005 and an indicator equal to 1 if the plan offers any type of asset allocation funds. Column (1) presents regression results without plan level controls, and column (2) presents results with the following plan-level controls: portion of female participants, average contributions, average account balance, average tenure, percentage of web-registered users, whether the plan offers a loan, company stock or brokerage account, and the size of the plan as proxied by the log number of participants. Errors are clustered at the plan level to further account for plan-level effects.

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1)</th>
<th></th>
<th>(2)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff</td>
<td>Std Err</td>
<td>Coeff</td>
<td>Std Err</td>
</tr>
<tr>
<td>HasAA</td>
<td>1.023***</td>
<td>0.377</td>
<td>1.084***</td>
<td>0.329</td>
</tr>
<tr>
<td>Contributions</td>
<td>0.367***</td>
<td>0.032</td>
<td>0.322***</td>
<td>0.032</td>
</tr>
<tr>
<td>HasAA*Contributions</td>
<td>-0.161***</td>
<td>0.058</td>
<td>-0.173***</td>
<td>0.050</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.994</td>
<td>0.213</td>
<td>-0.880***</td>
<td>0.351</td>
</tr>
<tr>
<td>Controls</td>
<td>(-)</td>
<td>(+)</td>
<td>(-)</td>
<td>(+)</td>
</tr>
<tr>
<td>Participants, plans</td>
<td>329,024</td>
<td>1,772</td>
<td>328,192</td>
<td>1,744</td>
</tr>
</tbody>
</table>
Table 3.4: Average Allocation to Equity by Age

This table displays the percentage of the portfolio invested in equities. Panel A includes participants that invest in asset allocation funds, though those investing in both risk-based and retirement date funds are excluded. Panel B includes all participants in plans that offer either risk-based or retirement date funds, but not both. Panel C is restricted to participants in these plans who have some exposure to equities, so participants without any equity exposure are excluded. The benchmark column refers to the average fraction allocated to equities in plans that do not offer asset allocation funds. *** indicates averages are significantly different at 1%; ** indicates at 5%; * indicates 10%.

<table>
<thead>
<tr>
<th>Panel A - Participants investing in asset allocation funds</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Investors in risk-based</td>
<td>Investors in retirement-date</td>
<td>Benchmark</td>
</tr>
<tr>
<td>Age</td>
<td>Equity (%)</td>
<td>Diff. from benchmark</td>
<td>Equity (%)</td>
</tr>
<tr>
<td>20-29</td>
<td>69.4</td>
<td>+0.5***</td>
<td>79.6</td>
</tr>
<tr>
<td>30-39</td>
<td>69.9</td>
<td>-0.4</td>
<td>73.7</td>
</tr>
<tr>
<td>40-49</td>
<td>68.5</td>
<td>+0.3*</td>
<td>62.8</td>
</tr>
<tr>
<td>50-59</td>
<td>65.0</td>
<td>+0.3</td>
<td>53.0</td>
</tr>
<tr>
<td>60-79</td>
<td>61.8</td>
<td>+2.6***</td>
<td>43.0</td>
</tr>
<tr>
<td>Young-old</td>
<td>7.6***</td>
<td></td>
<td>36.6***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B - All participants in plans that offer asset allocation funds</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Plans offer risk-based</td>
<td>Plans offer retirement date</td>
<td>Benchmark</td>
</tr>
<tr>
<td>Age</td>
<td>Equity (%)</td>
<td>Diff. from benchmark</td>
<td>Equity (%)</td>
</tr>
<tr>
<td>20-29</td>
<td>64.0</td>
<td>+0.8***</td>
<td>69.7</td>
</tr>
<tr>
<td>30-39</td>
<td>67.0</td>
<td>+0.5**</td>
<td>68.1</td>
</tr>
<tr>
<td>40-49</td>
<td>64.7</td>
<td>+0.7***</td>
<td>62.4</td>
</tr>
<tr>
<td>50-59</td>
<td>60.2</td>
<td>+0.7**</td>
<td>56.3</td>
</tr>
<tr>
<td>60-79</td>
<td>55.8</td>
<td>+2.7***</td>
<td>48.3</td>
</tr>
<tr>
<td>Young-old</td>
<td>8.2***</td>
<td></td>
<td>21.4***</td>
</tr>
</tbody>
</table>

Table continued on next page ...
Panel C - Participants in plans that offer asset allocation funds, conditional on owning equity

<table>
<thead>
<tr>
<th>Age</th>
<th>Plans offer risk-based Equity (%)</th>
<th>Diff. from benchmark (%)</th>
<th>Plans offer retirement date Equity (%)</th>
<th>Diff. from benchmark (%)</th>
<th>Benchmark Equity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-29</td>
<td>74.4</td>
<td>-1.8***</td>
<td>79.7</td>
<td>+3.5***</td>
<td>76.2</td>
</tr>
<tr>
<td>30-39</td>
<td>75.3</td>
<td>-2.3***</td>
<td>77.4</td>
<td>-0.2</td>
<td>77.6</td>
</tr>
<tr>
<td>40-49</td>
<td>73.8</td>
<td>-2.4***</td>
<td>72.2</td>
<td>-4.0***</td>
<td>76.2</td>
</tr>
<tr>
<td>50-59</td>
<td>70.6</td>
<td>-2.4***</td>
<td>66.4</td>
<td>-6.6***</td>
<td>73.0</td>
</tr>
<tr>
<td>60-79</td>
<td>68.4</td>
<td>-1.0</td>
<td>61.1</td>
<td>-8.3***</td>
<td>69.4</td>
</tr>
<tr>
<td>Young-old</td>
<td>6.0***</td>
<td></td>
<td>18.6***</td>
<td></td>
<td>6.8***</td>
</tr>
</tbody>
</table>

Table continued from previous page.
Table 3.5: Effect of Asset Allocation Funds on Equity Allocation

The table provides regression results for the following censored regression:

\[ PctEquity_{i,j} = \alpha + \beta \ast \left[ Age_{i,j} \cdot HasRB_j \mid Age_{i,j} \cdot HasRD_j \mid Age_{i,j} \right] + \epsilon_{i,j} \]

\( PctEquity_{i,j} \) is the percentage invested in equities by participant \( i \) in plan \( j \). \( HasRB_j \) and \( HasRD_j \) are indicators for whether plan \( j \) offers risk-based funds and retirement date funds, respectively. The regression model is estimated as a censored regression (lower bound = 0, upper bound = 100%). Column (1) presents regression results without plan level controls, and column (2) presents results with the following plan-level controls: portion of female participants, average contributions, average account balance, average tenure, percentage of web-registered users, whether the plan offers a loan, company stock or brokerage account, and the size of the plan as proxied by the log number of participants. Errors are clustered at the plan level to further account for plan-level effects. *** indicates coefficient is different than zero at 10%, ** is at 5%.

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1)</th>
<th></th>
<th>(2)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff</td>
<td>Std Err</td>
<td>Coeff</td>
<td>Std Err</td>
</tr>
<tr>
<td>Age</td>
<td>2.031***</td>
<td>0.429</td>
<td>1.852***</td>
<td>0.468</td>
</tr>
<tr>
<td>Age^2</td>
<td>-0.028***</td>
<td>0.005</td>
<td>-0.028***</td>
<td>0.006</td>
</tr>
<tr>
<td>HasRB</td>
<td>3.824</td>
<td>11.885</td>
<td>5.981</td>
<td>9.864</td>
</tr>
<tr>
<td>HasRB*Age</td>
<td>-0.175</td>
<td>0.550</td>
<td>-0.309</td>
<td>0.454</td>
</tr>
<tr>
<td>HasRB*Age^2</td>
<td>0.003</td>
<td>0.006</td>
<td>0.004</td>
<td>0.006</td>
</tr>
<tr>
<td>HasRD</td>
<td>31.886***</td>
<td>8.532</td>
<td>25.953***</td>
<td>8.582</td>
</tr>
<tr>
<td>HasRD*Age</td>
<td>-1.228***</td>
<td>0.416</td>
<td>-1.009***</td>
<td>0.391</td>
</tr>
<tr>
<td>HasRD*Age^2</td>
<td>0.011***</td>
<td>0.005</td>
<td>0.009**</td>
<td>0.005</td>
</tr>
<tr>
<td>Intercept</td>
<td>30.473***</td>
<td>8.937</td>
<td>35.945***</td>
<td>12.252</td>
</tr>
<tr>
<td>Controls</td>
<td>-</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>N Observations</td>
<td>329,024</td>
<td></td>
<td>328,192</td>
<td></td>
</tr>
</tbody>
</table>


Ball, R., Kothari, S., and Shanken, J., 1995. Problems in measuring portfo-


Brennan, M., and A. Subrahmanyam, 1995, Investment analysis and price for-


Fama, E.F., and French, K.R., 1992. The cross-section of expected stock returns,

Ferson, W.E., and Harvey, C.R., 1991. The variation of economic risk premiums,


Fidelity Investments, 2006. Building Futures: Volume VI.

Foster, F., and S. Viswanathan, 1996, Strategic trading when agents forecast the

Frederick, S., Loewenstein, G., O’Donoghue, T., 2002. Time discounting and

Gaver, J., and K. Gaver, 1993, Additional evidence on the association between the
investment opportunity set and corporate financing, dividend, and compensation

Gompers, P., and A. Metrick, 2001, Institutional investors and equity prices,

Goyal, A., and P. Santa Clara, 2003, Idiosyncratic risk matters!, *Journal of
Finance* 58, 975-1007.

Grossman, S. and J. Stiglitz, 1980, On the impossibility of informationally effi-

Grundy, B., and M. McNichols, 1989, Trade and the revelation of prices through

Hadlock, C., 1998, Ownership, liquidity, and investment, *Rand Journal of Eco-

151
nomic s 29, 487-508.


Iyengar, S., Kamenica, E., 2006. Choice overload and simplicity seeking. Unpub-
lished paper, University of Chicago,


Kotlikoff, L.J., Bernheim, B.D., 2001. Household financial planning and financial


Mankiw, N., Zeldes, S., 1991. The consumption of stockholders and nonstock-


