Essays on Index Premia and Demand Curves for Stocks

by

Antti Petäjistö

M.Sc., Engineering Physics, Helsinki University of Technology (1998)

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Management

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2003

© Antti Petäjistö. All rights reserved.

The author hereby grants to MIT permission to reproduce and
to distribute copies of this thesis document in whole or in part.

Signature of Author .................................................. Sloan School of Management
June 1, 2003

Certified by .......................................................... Stephen A. Ross
Franco Modigliani Professor of Financial Economics
Thesis Supervisor

Certified by .......................................................... Jiang Wang
Nanyang Technological University Professor of Finance
Thesis Supervisor

Accepted by .......................................................... Birger Wernerfelt
J.C. Penney Professor of Management Science
Ph.D. Program Chair
Essays on Index Premia and Demand Curves for Stocks

by

Antti Petäjistö

Submitted to the Sloan School of Management
on June 1, 2003, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Management

Abstract

This thesis consists of three chapters that investigate the index premium and its underlying economics both theoretically and empirically.

The first chapter presents our empirical findings about the index premium and its properties. First, we find that the index premia for both the S&P 500 and Russell 2000 have been growing over time, reaching levels of about 15% and 10%, respectively, in 2000. The premia arise somewhat gradually between the announcement and effective days and do not reverse at least in the next few weeks. Second, we find that the index premium is related to the idiosyncratic risk and market equity of a firm with economic and statistical significance. Third, we introduce a new concept that we label the index turnover cost, which represents a cost borne by index funds due to the index premium. We illustrate this cost and estimate it as 70-85 bp annually for the S&P 500 and 110-211 bp annually for the Russell 2000.

The second chapter develops the first theoretical explanation in the literature for downward-sloping demand curves. In traditional multi-asset models such as the CAPM, demand curves for stocks are almost perfectly horizontal, because a representative investor who is sufficiently risk-tolerant to hold the entire market portfolio has to be almost indifferent to idiosyncratic risk. We start with the basic CAPM setting, but we further assume that there is a fixed cost to actively managing a stock portfolio and that individuals pay the cost through an institution as a proportional fee. In equilibrium, the proportional fee can entirely determine the cross-sectional pricing of stocks, while the risk aversion of individual investors still determines the aggregate market risk premium. In contrast to any representative agent models, this allows demand curves for stocks to be sufficiently steep to have economic significance, also implying that stocks will be priced only approximately around their fundamental values. Our explanation can account for several empirically observed puzzles such as the magnitude of the S&P 500 index premium.

The third chapter focuses on index investors for whom the index premium creates a recurring cost: as the index is updated, they need to buy stocks with the premium and sell stocks without the premium. Different index rules can produce different index premia due to the different frequency and criteria of updating. We build a model to investigate the behavior of the index turnover cost and the portfolio performance of a mechanical index fund under a market-cap rule, an exogenous random rule, and a deterministic rule. We find that the rational anticipation of future index composition reflected in prices today eliminates any first-order differences in index fund performance across the three index rules. As the index investors become a large part of the market, the non-index investors become less diversified, and this induces hedging motives.
which hurt the index investors especially under a market-cap rule.

Thesis Supervisor: Stephen A. Ross
Title: Franco Modigliani Professor of Financial Economics

Thesis Supervisor: Jiang Wang
Title: Nanyang Technological University Professor of Finance
Acknowledgments

I want to express my deep gratitude to the members of my dissertation committee: Steve Ross, Dimitri Vayanos, and Jiang Wang. Their insightful comments and questions forced me to think more deeply about my own research ideas, and as a result their feedback accounted for a significant part of my learning in the Ph.D. program. In particular, Steve's broad understanding of finance, Dimitri's attention to the specifics of my economic arguments, and Jiang's comments on both types of issues have all been extremely valuable to me.

I am also very grateful to Bengt Holmström for his insights on the general philosophy of economics research. He has illuminated many crucial points to me that I might have spent a career finding out for myself. In addition, Tuomo Vuolteenaho has helped me countless times on a wide range of finance-related questions. A large part of my knowledge in empirical finance was accumulated during a research assistantship with Andrew Lo and Jiang Wang. I also owe a lot to my fellow finance Ph.D. students for providing an environment that was not only intellectually stimulating but also comfortable and friendly.

For financial support, I am deeply indebted to the Finnish Cultural Foundation, the Osk. Huttunen Foundation, the Emil Aaltonen Foundation, and the Yrjö Jahnsson Foundation. Their generosity allowed me to focus all my energy on learning and doing research instead of worrying about my personal finances.

Finally, I want to thank my sweet girlfriend Reeta for her constant support, encouragement, and advice. Her genuine and warm smile always has a miraculous ability to make me feel happy.

I dedicate this thesis to all those who promote freedom and its immediate corollary, the free market, all around the world.
Contents

1 The Index Premium and Its Implications for Index Funds 9
   1.1 Introduction ............................................. 9
   1.2 S&P 500 Index Premium .................................. 11
       1.2.1 Background ........................................... 11
       1.2.2 Index Premium ........................................ 12
   1.3 Russell 2000 Index Premium ............................... 17
       1.3.1 Background ........................................... 17
       1.3.2 Index Premium ........................................ 18
   1.4 Cross-Sectional Dependence of the Premium ............. 20
       1.4.1 Russell 2000 .......................................... 21
       1.4.2 S&P 500 ............................................... 23
   1.5 Index Turnover Cost ...................................... 25
       1.5.1 S&P 500 ............................................... 25
       1.5.2 Russell 2000 .......................................... 28
   1.6 Conclusions ............................................... 30
   1.7 References ............................................... 31
   1.8 Tables .................................................... 33
   1.9 Figures ................................................... 44

2 What Makes Demand Curves for Stocks Slope Down? 54
   2.1 Introduction ............................................. 54
   2.2 The Puzzle: Theory and Empirical Evidence ............. 58
2.2.1 Traditional Arguments ........................................... 58
2.2.2 A Simple CAPM Calibration ................................. 59
2.2.3 Empirical Evidence ........................................... 60
2.2.4 Suggested Hypotheses ........................................ 61
2.3 An Explanation with Financial Intermediaries ................. 64
  2.3.1 Motivation .................................................. 64
  2.3.2 The Model .................................................. 65
  2.3.3 Analysis of Equilibrium .................................. 72
  2.3.4 Empirical Implications .................................... 75
2.4 Empirical Tests .................................................. 77
2.5 Interpretations and Further Discussion ......................... 78
  2.5.1 Optimality of Active Managers’ Policy .................. 78
  2.5.2 Relationship to Grossman and Stiglitz (1980) ......... 79
  2.5.3 Institutions Make Prices More Efficient ................. 81
  2.5.4 Transaction Costs ......................................... 82
  2.5.5 CRRA Utility .............................................. 83
  2.5.6 Interpretation of Long-Short Positions .................. 83
2.6 Conclusions ....................................................... 84
2.7 References ....................................................... 86
2.8 Appendix: A More Elaborate Model ............................ 91
  2.8.1 Motivation .................................................. 91
  2.8.2 The Model .................................................. 91
  2.8.3 Analysis of Equilibrium .................................. 95
2.9 Appendix: Derivation of Formulas .............................. 98
  2.9.1 CAPM Benchmark .......................................... 98
  2.9.2 Active Manager ............................................ 98
  2.9.3 Analogy with a CARA Investor ............................ 99
  2.9.4 Equilibrium ............................................... 100
2.10 Figures .......................................................... 103
3 Selection of an Optimal Index Rule for an Index Fund

3.1 Introduction ............................................. 105
3.2 Behavior of Index Funds ................................ 109
3.3 The Model ................................................ 110
  3.3.1 Motivation ........................................... 110
  3.3.2 Model Setup .......................................... 110
  3.3.3 Equilibrium ........................................... 112
3.4 Analysis of Equilibrium ................................. 116
  3.4.1 Market-Cap Rule ..................................... 116
  3.4.2 Exogenous Index Rules ............................... 121
  3.4.3 Portfolio Performance ............................... 123
  3.4.4 Robustness of Results ............................... 126
3.5 Extension to Continuous Time ........................... 128
  3.5.1 Motivation ........................................... 128
  3.5.2 Model Setup .......................................... 128
  3.5.3 Equilibrium ........................................... 129
  3.5.4 Analysis of Equilibrium ............................... 132
  3.5.5 Portfolio Performance ............................... 133
3.6 Interpretations and Extensions ........................... 133
  3.6.1 Interpretation of Results ........................... 133
  3.6.2 Index Effect for a Large Number of Stocks .......... 135
3.7 Conclusions ............................................... 138
3.8 References ................................................. 140
3.9 Appendix: Formulas in Discrete Time ...................... 142
  3.9.1 Equilibrium Prices at $t = 2$ ...................... 142
  3.9.2 Equilibrium Prices at $t = 1$ ...................... 143
  3.9.3 Equilibrium Prices under a Market-Cap Rule ....... 145
  3.9.4 Equilibrium Prices under an Exogenous Index Rule .. 147
3.10 Appendix: Formulas in Continuous Time .................. 148
  3.10.1 Equilibrium Prices under a Market-Cap Rule ....... 148
Chapter 1

The Index Premium and Its Implications for Index Funds

1.1 Introduction

According to neoclassical finance, the price of a stock is given by its expected future cash flows discounted by their systematic risk. If the price were to deviate from this fundamental value, aggressive arbitrageurs would quickly step in and eliminate the mispricing. In a world with thousands of stocks, an individual stock will be an almost negligible part of an investor’s portfolio, which makes it easy for arbitrageurs to buy or sell relatively large quantities of an individual stock. Hence, when there is a clearly uninformed supply or demand shock for a stock, the demand by arbitrageurs should be almost perfectly elastic, and that is why we expect demand curves for stocks to be almost perfectly horizontal.

This traditional view, held by many since the advent of the CAPM in the 1960s and widely publicized in textbooks and articles, has been challenged by mounting empirical evidence for downward-sloping demand curves. In particular, the increasing popularity of indexing has provided researchers with a relatively clean experiment to test for the slope of the demand curve. When index funds mechanically buy a nontrivial fraction of all the stocks in an index, the demand by these funds represents an uninformed demand shock whenever a stock is added to or deleted from an index. The index-induced price effect can then be used to infer the slope of the demand curve.
The early evidence for the S&P 500 Index (e.g., a 3% price impact documented by Shleifer (1986), and a 5-7% price impact documented by Lynch and Mendenhall (1997) a decade later) has recently been complemented by similar evidence from international markets, usually with somewhat different index selection rules. Now in 2003, we venture to claim that the evidence for downward-sloping demand curves is overwhelming.

In this chapter, we first want to find out what the current state of the S&P 500 premium is. Indexing has surged in popularity since the last systematic and well-known study by Lynch and Mendenhall (1997) whose data ended in 1995, so it is natural to ask whether the current higher fraction of indexing has materially affected the size of the index premium. We conduct a year-by-year analysis of the premium to identify a pattern across the years. Also within each year, we investigate the evolution of the premium day-by-day around the index announcement and the effective day of a change. This can give us a better idea of how efficiently the market is responding to this kind of new information. We also look at the long-run evolution of the premium to identify a possible reversal.

In addition to the S&P 500, we run similar tests for the Russell 2000. So far there is virtually no academic research about this index, as researchers have focused on the S&P 500 or international stock indices. Nevertheless, the Russell 2000 does provide a very large sample, with hundreds of additions and deletions every year, which of course makes it very useful for research purposes. We also estimate the slope of the demand curve for both the S&P 500 and Russell 2000.

The second set of new results in this chapter comes from cross-sectional regressions of the index premium on explanatory variables. We pick idiosyncratic risk and market equity as our theoretical model in Chapter 2 suggests that.

Finally, we introduce a new concept that we label the “index turnover cost.” This reflect the recurring costs that a mechanical indexer will have to pay for always buying stocks with the index premium and selling them without the premium. The cost is measured against what we label an “index-neutral” strategy, which consists of holding a portfolio with essentially identical characteristics but not being mechanically tied to holding the index all the time. This is a cost that index investors should certainly care about. We pick simple examples to illustrate this cost for both the S&P 500 and Russell 2000, and we estimate the size of this cost from data.
We find that the index premia for both the S&P 500 and Russell 2000 have been growing over the 1990s with the popularity of indexing, reaching peaks of 15.1% and 9.5% for additions in 2000. The premium does not reverse in the next few weeks, but beyond that it is hard to infer much – sometimes the premium reverses, and sometimes it keeps increasing. It is also surprising that there is a very clear and economically significant drift for both indices between the announcement and effective days, and this drift persists up to 2000, the last year in our data. The index premium is related to idiosyncratic risk with statistical and economic significance, and it also seems to be related to market equity, albeit with not quite the same level of significance.

The annual index turnover cost in the most recent year turns out to be about 70 bp for the S&P 500 and 110 bp for the Russell 2000%. If the index-induced price effects are not permanent, the cost would be even greater. This is something that index investors might want to think about e.g. when choosing between an S&P 500 fund and a Wilshire 5000 fund.

This chapter proceeds as follows. Section 2 investigates the S&P 500 index premium. Section 3 does the same for the Russell 2000. The cross-sectional regressions for both indices are presented in section 4. Section 5 illustrates and estimates the index turnover cost for both indices. Section 6 concludes.

1.2 S&P 500 Index Premium

1.2.1 Background

The Standard and Poor's 500 Index consists of a sample of 500 firms intended to be representative of the U.S. economy. At the end of 2000, the aggregate market value of the firms in the index was roughly $11 trillion, accounting for about 76% of the total U.S. stock market capitalization. If we compare the S&P 500 firms with all U.S. firms ranked by market capitalization, we find about 74% of the index firms within the top 500, about 94% within the top 1,000, and about 99% within the top 1,500 firms. These percentage shares account for 97%, 99.8%, and 99.98% of the market value of the S&P 500. Stocks within the index are value-weighted based on the total stock market capitalization of each index firm.

The index is selected by the S&P index selection committee. Selection is based on criteria
such as market capitalization, industry representation, liquidity, trading volume, and financial soundness. The committee makes its decisions behind closed doors according to somewhat broad guidelines, and it has not announced a mechanical rule for its decisions. Market capitalization seems to be a very important criterion, but from an investor's point of view significant residual uncertainty remains about index changes. Index changes are always prompted by deletions. Once the committee makes a decision about a deletion, it will select the replacement from a pool of candidate stocks shortlisted for addition.

After October 1989, S&P has attempted to announce index changes usually about five trading days before they become effective. Prior to October 1989, index changes would become effective immediately after announcement. Both the actual changes and announcements are made after the close of trading. Announcements occur at apparently random times throughout the year, usually only one but almost always less than five stocks at a time.

There is a large number of mutual funds and pension funds that mechanically track the S&P 500 index. Indexing as an investment style began to grow dramatically since the 1980s, and the trend has continued all the way through the 1990s (Morck and Yang (2001)). At the end of 2000, S&P estimated that more than $1 trillion was directly indexed to the S&P 500, accounting for 10% of the market value of the stocks in the index.¹ In the long run, according to a chief executive at the indexing giant Barclays,² 50% of the stock market is "not an unreasonable target" for index funds.

1.2.2 Index Premium

Whenever a stock is added to the S&P 500 index, about 10% of its shares outstanding will be bought by mechanical indexers. Most significant indexers actually hold all 500 underlying stocks and try to update their portfolios as close as possible to the actual index change in order to minimize their tracking error relative to the index (Blume and Edelen (2001)). This creates an economically large demand shock by mechanical indexers.

According to neoclassical finance, investors hold well-diversified portfolios and can easily absorb such an uninformed demand shock for one stock. Thus the demand curve for a stock

¹Higher estimates have been presented (e.g. Beneish and Whaley (1990) already suggest 10% at a time when indexing was not yet as popular as it is now), but today the most commonly quoted estimate is 10%.

by the representative investor should be almost perfectly horizontal, and even a large demand shock around an index change should not have an economically significant effect on the price of the stock. Conversely, if the demand curve for a stock does have a nontrivial negative slope for one reason or another, even an easily identifiable uninformed demand shock will have a meaningful price impact.

An investigation of the price impact of S&P 500 index addition thus provides a relatively clean test of the slope of the demand curve for a stock. Consequently, it will also serve as a test of our neoclassical asset pricing theories, namely the CAPM and in an informal sense also the APT.

**Estimation Procedure**

We use data for S&P 500 changes in 1980-2000. After October 1989, we distinguish between the announcement day and effective day of a change. We combine this with the CRSP daily stock files to perform our analysis.

For each event stock (CRSP permno), we require valid CRSP return data 15 trading days before the announcement day and 15 trading days after the effective day. This eliminates firms that undergo M&A activity such as being spun off or acquired by another firm.\(^3\) It also eliminates firms that were delisted from their exchanges only a few days after their deletion from the index, which may occur for firms experiencing sudden financial distress. In the remaining sample there are more than twice as many additions as deletions, where the difference is accounted for by mergers between two index firms.

We define the effective day of the index change as trading day zero in event time, so the index is updated using the closing prices of trading day zero. The announcement usually occurs after close on trading day \(-5\). We define the cumulative abnormal return (CAR) on a stock as the difference between the cumulative stock return and the cumulative return on the CRSP value-weighted market index, expressed as a percentage of the cumulative gross return on the market index. Note that since index changes depend on past returns, this induces a selection bias to the alpha estimates of the market model, so the standard market model is not

---

\(^3\)Of course this is an imperfect filter to screen out all M&A-related events from the dataset, but it seems implausible that the few remaining cases (spinoffs of index firms where both new firms were deleted from the index) would systematically bias our results.
applicable here and that is why we use market-adjusted returns instead. We normalize the
CAR to zero at trading day $-20$. This allows us to identify a possible pre-announcement drift
due to information leakage or other anticipation of the index announcement.

Since the index premium is not likely to be constant across our sample period of 20 years,
especially given the increasing relative size of index funds, we conduct our analysis separately
for each year. Each year we form an equally weighted portfolio of additions at trading day $-20$
(with their effective days that year), and another portfolio of deletions at trading day $-20$. We
then look at the CARs of those portfolios, i.e. the cumulative buy-and-hold abnormal returns
on the portfolios of event stocks.

To compute standard errors, we estimate for each stock the volatility of market-adjusted
returns using a 6-month period of daily data ending one week before the announcement day.
Since index changes occur throughout the year, we assume the abnormal returns across stocks
are uncorrelated when we estimate the volatility of the event stock portfolio. In reality there
is of course some overlap in calendar time, especially since occasionally 2-3 stocks are replaced
at the same time. As our market adjustment is unlikely to eliminate all systematic risk, our
standard errors are likely to be somewhat understated.

As we are interested in the price effects around both the announcement day and the effective
day, we want to align both days for all event stocks. While the most common difference between
the announcement and effective days is 5 trading days, this difference may even vary from zero
to about one month. Hence, we form two samples. The first one consists of stocks where
the difference is at least 2 days. If the difference is 2-4 days, we “stretch” the returns to
cover 5 days in event time. If the difference is greater than 5 days, we shrink the interval to
5 days. This allows us to align the CAR at the close of trading following the announcement
day and the CAR at the close of trading on effective day for all stocks in the sample. The
second sample consists of the first sample (with the aligned event days) plus the stocks where
the announcement was 0-1 days before the change, where the event time of the latter group is
not altered.
Results

Table 1.1 shows the CARs in 1990-2000 at trading day zero, i.e. over a period of about one month, along with the associated standard errors. The table also shows the number of qualifying additions and deletions each year. The number of index changes is greater towards the end of the sample, reaching 50 additions and 23 deletions in 2000, which makes these later estimates more reliable. It also seems plausible that the index premium is greater at the end when indexers comprise a larger part of the market. The CAR in 2000 is 15.1% for additions and −18.9% for deletions.

Figure 1-1 describes the evolution of the CARs around the event. The additions are flat until the last 5 trading days before announcement, when they accumulate a 2-3% premium. As expected, the biggest jump occurs right after announcement. Nevertheless, the announcement return remains small enough to leave an economically significant 1% daily drift between the announcement and effective days. Given the publicity surrounding S&P index changes and the associated price premium, it is indeed surprising that such a drift can still persist and has not been arbitraged away.\footnote{In fact there is anecdotal evidence of hedge funds trading on this drift but apparently not enough to eliminate it.} The CAR peaks at the effective day and then experiences approximately a 2% reversal in the next few days. No further reversal occurs at least in the next 2 weeks.

The deletions behave somewhat similarly, except that there seems to be a greater anticipation of the announcement. Part of the negative pre-announcement return could also be due to the S&P deleting stocks that just experienced sharp losses in market value and thus creating a selection bias. Since all index changes are prompted by deletions, and since extreme negative returns are more common than extreme positive returns, this selection bias is unlikely to affect the CAR for additions.

Figure 1-2 tells a similar story about CARs in 1999. The only differences are the lack of any short-term reversal and an apparently greater anticipation of deletions ahead of the announcement, although the standard errors are too wide to allow us to draw any sharp new conclusions.

The long-horizon returns are described by Figures 1-3 and 1-4. In 2000, the additions seem
to reverse fully in about 8 months, while the deletions reverse fully in a month and actually gain a temporary premium of 20% only 3 months after the event. Since deletion from the S&P 500 per se cannot conceivably be good news for a stock, this peculiar behavior of the CARs simply suggests that our sample is so small that it does not have much power to determine the long-run price effects. This view is supported by the 1999 price effects which keep increasing after the event, reaching 30% for additions 9 months after the event and -25% for deletions already one month after the event. Data from earlier years are similarly inconclusive. Hence, as in the earlier empirical literature, we are unable to conclude how much of the index-induced price effect reverses in the long run or whether there is a permanent price premium for all stocks as long as they remain in the index. The premium does seem at least semi-permanent, since no more than a small fraction of the price effect reverses in the first one month.

The evolution of the index premium from 1980 to 2000 is illustrated in Table 1.2 and Figure 1-5. In the 1980s, the premium for additions seems fairly stable at 2-5%. In 1990-1992, the premium estimates jump, possibly because of the October 1989 rule change or simply because of the very small sample size these years. After 1992, the premium increases more or less steadily over the decade up to 2000. Our numbers are consistent with the earlier literature: Shleifer (1986) finds an addition premium of 3% for the 1976-1983 period, while Lynch and Mendenhall (1997) find an addition premium of about 5-7% for 1990-1995.

The increase in the premium in the 1990s coincides with an economically significant increase in the share of mechanical indexers over this period. Morck and Yang (2001) point out that the value (in 1982 dollars) of Vanguard 500 was $0.11 billion in 1982, $1.66 billion in 1990, and $32.3 billion in 1997, representing percentage shares of the S&P 500 index of 0.01%, 0.10%, and 0.69%, respectively. Since the Vanguard 500 is the oldest and largest S&P 500 index fund, its growth seems like a reasonable proxy for the growth rate of the index fund industry over this period. The current total share of mechanical indexers is most commonly estimated to be around 10% of the index.

Estimating the price elasticity of demand for a stock, we find $\frac{\Delta Q}{\Delta P} = -0.7$ using the most recent numbers from the year 2000. This is relatively close to the unit price elasticity of demand estimated by Shleifer (1986). The most significant source of noise in the estimate comes from the share of indexers: while it is relatively easy to estimate the share of purely
mechanical indexers, a very large number of funds are benchmarked against the index which creates additional long-run demand for index stocks. However, it seems unlikely that any such funds would regularly rush to trade on index changes within a few days of the effective day.

1.3 Russell 2000 Index Premium

1.3.1 Background

Another widely tracked stock market index is the Russell 2000. It measures the performance of 2,000 small-cap stocks. Unlike the S&P indices which are selected by a committee, Russell selects its indices mechanically based on market capitalization. The Russell 3000 index covers the largest 3,000 U.S. stocks, and it consists of the Russell 1000 index, which covers the largest 1,000, and the Russell 2000, which covers the remaining 2,000 stocks. The lower cutoff of the Russell 2000 has recently been around $150 million and the upper cutoff around $1.5 billion. The Russell 2000 is the most well-known and closely followed of all the Russell indices. Its market capitalization in May 2000 was $1.2 trillion. An estimated $180 billion is tied to the Russell indices with more than $25 billion tracking the Russell 2000.\(^5\) We can thus estimate that about 2% of the Russell 2000 is held by passive indexers. Another $800 billion is benchmarked against the Russell indices, creating additional long-term demand for the index stocks.

The Russell indices are updated once a year based on market capitalization at the market close on May 31. The changes become effective one month later on July 1. Indexers typically update their portfolios at the end of June in order to minimize their tracking error. Unlike the S&P, Russell determines index weights based on its definition of the public float of a firm and not its total number of shares outstanding. This is done in an effort to keep the index easily tradeable, which is a concern especially for a small-stock index.

A clear index effect exists for the Russell 2000, although it has not received as much attention as the S&P 500 effect. In 2000, 684 stocks were added to the Russell 2000. During the month between the announcement date and the effective date, these stocks rose 38% on average. Since index membership is based on market capitalization alone, the index changes are relatively easy to predict a few weeks before the new index stocks are determined. In 2001, Goldman Sachs

estimates that during the 8 weeks preceding May 31, the stocks expected to be added to the index outperformed the index by 22%. Barclays says it started its own analysis of the 2001 index changes in February, more than three months before May 31.\(^6\)

1.3.2 Index Premium

Estimation Procedure

We use data from 1987 to 2000, obtained from Frank Russell Co. Prior to 1987, the Russell indices were updated quarterly instead of annually. We also do not wish to go too far back in the data, since there is no reason to believe that the share of indexers and their behavior around index changes would stay constant throughout.

Because all index changes in a year occur exactly at the same time, we can expect significant cross-correlation for stocks. Hence, we want to control also for systematic non-market factor exposure such as variations in market equity. In anticipation of our cross-sectional regressions in Section 1.4, we also control for differences in idiosyncratic risk (defined as the residual volatility from the Fama-French 3-factor model). We defer a more detailed discussion of the benchmark portfolios to Section 1.4. Based on the market capitalization of the stock on April 30 and its idiosyncratic volatility estimated from the preceding 6 months, we allocate each event stock into its matching portfolio in a \(10 \times 5\) grid of market equity and idiosyncratic risk. We define the abnormal return on a stock as the difference between the cumulative stock return and the cumulative return on the corresponding buy-and-hold control portfolio, expressed as a percentage of the cumulative gross return on the control portfolio. For abnormal returns in June, we use the estimation period ending on April 30 to determine the benchmark portfolio compositions and to assign each event stock to its corresponding portfolio, and for abnormal returns in July, we use the period ending on May 31, and so on.

We start computing the CARs after the close on May 31. This is effectively the announcement day of the event, since market participants can use the closing prices on May 31 to determine the future index composition with virtually no uncertainty.\(^7\) Since some additions and deletions can be predicted earlier than May 31, part of the index-induced price effect is

---

\(^6\)The Wall Street Journal, 5/31/01.
\(^7\)Russell usually makes the official announcement 1-2 weeks later.
likely to occur before May 31. Yet we focus on returns after May 31, primarily because we
do not want to be subject to a forward-looking selection bias but also simply because over the
years most of the abnormal returns seem to occur in June, in spite of the anecdotal evidence
of some investors predicting index changes well ahead of time.

As the Russell 2000 has both a lower cutoff and an upper cutoff, we distinguish between four
kinds of event stocks: additions from below and above, and deletions from below and above.
Stocks crossing the upper cutoff are bought and sold not only by Russell 2000 indexers, but the
opposite side of the trades is taken by Russell 1000 indexers. These opposite effects are likely
to confound the price effects, and hence the additions and deletions around the lower cutoff
seem to be a cleaner sample for our test of demand curves.

We obtain the standard errors by computing daily abnormal returns for the event portfolios
(the addition portfolios and the deletion portfolios) from March 1 to May 31, and then using
the volatility of these abnormal returns for the event portfolio in June.

Results

Table 1.3 and Figure 1-6 show the cumulative abnormal returns for Russell 2000 additions and
deletions that cross the lower cutoff of the index. The CARs for additions are fairly modest,
usually around 2-4% until the later part of the 1990s. In the last four years, 1997-2000, the
addition premium seems to have increased to almost 10%. The CARs for deletions are slightly
greater and negative, but the increasing pattern for them is less evident. For both additions
and deletions, we see a small but peculiar peak in CARs around 1987-1988. This might reflect
the rule change in 1987 and a slow adjustment by investors to the new annual reconstitution of
the index.

The qualifying sample (with the data availability requirement) still has a large number of
event stocks: 442 additions and 318 deletions in 2000. This is a far larger sample than for the
S&P 500 and good news for the statistical power of our tests. The downside is the concentration
of the event in calendar time and the subsequent cross-correlation between stocks which reduces
the benefits of the large sample.

The sample of additions and deletions crossing the upper cutoff (Table 1.4) does not exhibit
any immediately clear patterns. The CARs of additions seem to be zero, and while the CARs
for deletions may appear slightly negative on average, they could also very well be zero. This is what we would expect if the share of indexers in Russell 1000 is comparable to the share of indexers in Russell 2000.

Figure 1-7 shows the evolution of the CARs for additions and deletions from below in 2000. The premium seems to accrue rather steadily between May 31 and June 30, with a peak one trading day before the effective day. There is a 2-3% reversal from the peak, but the rest of the premium stays for at least two weeks after the event. The pattern is similar for 1999 (Figure 1-8), except that the peak occurs one day later on effective day and there is no reversal.

Figures 1-9 and 1-10 illustrate the long-run price effects for additions and deletions. In 2000, most of the announcement CAR reverses in 6 months, but in 1999 it doubles in 6 months. Deletions also partially reverse in 2000 but continue to go down in 1999. These graphs again do not allow us to conclude much about whether the price effect is permanent. We can only say that the premium stays for at least a few weeks. The earlier years yield similarly inconclusive results.

Using the latest numbers for the elasticity of demand, we plug in 2% for $\frac{\Delta q}{q}$ and 9.5% for $\frac{\Delta P}{P}$, which gives us an elasticity of $-0.2$. The elasticity of demand therefore seems to be more than 3 times as great for the S&P 500 index changes as for the Russell 2000 index changes. Again, the same caveats apply for the percentage share of indexers.

1.4 Cross-Sectional Dependence of the Premium

Is there any variation in the index premium across stocks? Can we identify any characteristics that would be associated with a higher elasticity of demand for a stock? Here we investigate two candidate variables: the market value of equity, and the idiosyncratic risk of a stock.

Market equity could be associated with the elasticity of demand e.g. if investors are segmented into different subsets of the market depending on the market value of the firms they trade and if these different investor segments behave somewhat differently. The idiosyncratic risk of a stock could also matter, as a greater idiosyncratic volatility makes it riskier to take positions even against clearly uninformed traders. However, note that the impact of idiosyncratic risk in a CAPM world should essentially be zero for a well-diversified investor, and in such a
world we should not be able to empirically distinguish any economically meaningful effect.

1.4.1 Russell 2000

Methodology

We choose our event window as June 1 to June 30. Hence, we test if the abnormal return in June on a stock added to the index from below is positively related to the idiosyncratic risk or market equity of the stock. We use data from 1987-2000 when the index has been reconstituted annually.

To estimate idiosyncratic risk, we use 6 months of daily data from CRSP from November 1 to April 30. We require a minimum of 2 months of valid return observations. We regress the stock’s daily excess return on the three factors of Fama and French.\(^8\) We define idiosyncratic risk as the root mean squared error of this regression.\(^9\) We also take the market equity of every firm on April 30 in order to obtain a value that is not affected by the anticipation of the index event.

Since it is possible that the level of idiosyncratic risk is also related to the cross-correlations of stocks, e.g. stocks with high idiosyncratic risk tend to move together, we need to control for this comovement of stocks with similar idiosyncratic risk. The market equity of a firm can also plausibly be associated with the slope of the demand curve, so we control for that as well.

Hence, we form a \(10 \times 5\) matrix of control portfolios based on market equity and idiosyncratic risk. We pick all stocks in CRSP representing ordinary common shares of U.S. firms on April 30, and we sort them into 10 deciles based on the Fama-French breakpoints for market equity that month. Having estimated the idiosyncratic risk of each stock as described before, we then subdivide each market equity decile into quintiles based on idiosyncratic risk. The procedure is similar to the one used by Fama and French (1992) for market equity and beta. We perform a sequential sort rather than an independent sort because idiosyncratic risk and market equity have a high negative correlation (about \(-0.5\) in a typical cross-section for idiosyncratic risk and log of market equity), so an independent sort would tend to cluster the stocks in the cells around the diagonal. After all, our purpose is to distinguish between levels of idiosyncratic risk

\(^8\)http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
\(^9\)We also ran all tests with the market model and obtained very similar results.
within each size decile, and except for the bottom and top size deciles the correlation between the two within a size decile is relatively small (generally between $-0.1$ and $0$). For a similar reason, we use all stocks (and not just NYSE stocks) for idiosyncratic risk breakpoints, since a large fraction of our event stocks are not NYSE stocks. Some sample statistics for the control portfolios are in the appendix.

We then compute the return on each control portfolio for each trading day in June. On May 31, we set the portfolio weights based on market capitalization on April 30. We calculate the buy-and-hold return on this portfolio for each day in June, assuming that when a stock no longer has valid CRSP return data we reinvest that wealth in the remaining portfolio. Note that this approach avoids any biases that might arise from the bid-ask bounce for small stocks in the presence of continuous rebalancing of portfolios.

Having determined the breakpoints for market equity and idiosyncratic risk, we can assign each event stock to its corresponding cell in the $10 \times 5$ matrix. We then define the cumulative abnormal return on a stock in June as the difference between the cumulative stock return and the corresponding cumulative control portfolio return.

Since all index changes each year occur at the same time, the abnormal returns are likely to exhibit significant cross-correlation. To get around this issue, we run a Fama-MacBeth regression for the 14 annual cross-sections of data covering the years 1987-2000. In other words, using all stocks within a year, we regress the abnormal stock return on idiosyncratic risk and log of market equity. Then we compute the time-series means of these regression coefficients and the standard errors for the means.

We use both additions and deletions for the analysis. Since the index-induced demand shock for deletions has a negative sign, we simply multiply the CARs for deletions by $-1$ and then lump them together with the additions. This allows us to use a relatively large sample in our regressions.

**Results**

The univariate Fama-MacBeth regression produces a significant $t$-statistic of 3.67 for the coefficient of idiosyncratic risk (Table 1.10). The $t$-statistic drops to 2.47 when we add log market equity into the regression. However, market equity does not appear to be statistically signifi-
cant either in the univariate or bivariate regression ($t = -1.77$ and $-1.47$, respectively). Since idiosyncratic risk is negatively correlated with market equity, it is perhaps not so surprising to find a negative coefficient for market equity, as our three-factor residuals are certainly not perfect proxies for idiosyncratic risk.

Economically the coefficient of idiosyncratic risk implies that an increase of 10% in annual idiosyncratic volatility would increase the price impact of Russell 2000 addition by 0.4%, or about 10% of the average price impact over this period. Given that the index premium at the end of the sample is about twice the sample average, this could translate to an almost 1% difference in CAR due to a 10% difference in idiosyncratic volatility. This is not a trivial magnitude, especially as our coefficient estimate is likely to be biased down due to the noisy measurement of idiosyncratic risk.

### 1.4.2 S&P 500

We then proceed to run a similar test with S&P 500 data. Here it should be pointed out that a similar test, albeit with a very different implementation, has already been carried out with S&P 500 data for the 1976-1989 period by Wurgler and Zhuravskaya (2002). The authors find evidence of a link between idiosyncratic risk and the price impact around index addition. However, their tests do not touch the more recent data since the index rule change in October 1989, so we want to confirm the result with data from the 1990s, especially as the fraction of mechanical indexers has grown so much since the 1976-1989 period. We also want to run the test with our methodology of controlling for idiosyncratic risk in the abnormal returns. Furthermore, this allows us to link the S&P 500 results to the Russell 2000 results.

#### Methodology

We want to test whether idiosyncratic risk and market equity are related to the price impact of an index-induced uninformed demand shock. We again define the price impact as the cumulative abnormal return from 15 trading days before the announcement to the close on the effective day.

We try to follow our procedure for the Russell 2000 as closely as possible. We take the last end-of-month market value of equity three weeks before the announcement day (i.e. about
a month before the effective day) of the index change. We choose the 6-month period ending
with the measurement day for market equity, and we estimate idiosyncratic risk in this period
as the root mean squared error of a regression of daily excess returns on the three factors of
Fama and French. All additions in the sample have at least 2 months of return observations
in the estimation period.

We use the same $10 \times 5$ matrix of control portfolios as before. Naturally the S&P 500
additions end up in the large-cap cells in the matrix, while the deletions can also be found
in small-cap cells. The cumulative abnormal return on a stock in the event window is then
the difference between its own cumulative return and the cumulative return of its benchmark
portfolio (matched on market equity and idiosyncratic risk).

For the S&P 500 the event windows are more or less randomly distributed throughout
the year, so cross-correlations of abnormal returns are not likely to be an important issue
here. Hence, each observation represents an independent data point and we can regress all the
observations on the explanatory variables in one cross-section. We perform such a regression,
and we use White's heteroskedasticity-consistent standard errors for inference. We also plug in
a dummy variable for each year to account for the increasing time trend in the index premium.

Results

Table 1.12 shows the regression results. Idiosyncratic risk turns out to be statistically significant
both in the univariate regression ($t = 4.50$) and in the bivariate regression ($t = 3.47$). Market
equity also turns out to be statistically significant ($t = -3.42$ and $-2.06$). When the year
dummies are not present, the coefficient of market equity is somewhat greater as the increasing
size of event stocks picks up some of the increasing time trend in the CARs over the years.

The coefficient of idiosyncratic risk is about 0.2, meaning that an increase in annualized
idiosyncratic volatility of 10% would increase the abnormal return around index changes by 2
percentage points. This certainly has economic significance, especially as the estimates are
still likely to be biased down due to measurement error.
1.5 Index Turnover Cost

For an index fund that mechanically tracks an index, the index premium represents a recurring cost. Whenever a stock is added to the index, the index fund has to buy it at a price which includes the premium. Whenever a stock is deleted from the index, the index fund has to sell it at a price which does not include the premium. In contrast, an unconstrained investor can buy non-index stocks without the premium, and if such a stock is later added to the index, the investor will actually earn the index premium for himself. The extra cost due to mechanical indexing relative to an index-neutral strategy is what we label the index turnover cost.

Index funds themselves probably do not care about the index turnover cost per se, because the fund managers are not penalized when the index itself suffers from the same cost. However, the people who invest in index funds certainly should care about this cost, as they could potentially invest in similar index-neutral portfolios and avoid the cost.

In this section we illustrate the index turnover cost and estimate its size for the two most commonly tracked indices, the S&P 500 and Russell 2000.

1.5.1 S&P 500

Source of the Cost

To illustrate the index turnover cost for the S&P 500, let us think about a simplified example.

Let the index consist of randomly selected stocks accounting for a fraction $s$ of a commonly known pool of large stocks. These index stocks have a price premium $p$ (relative to what the price would be without index funds). Additions to the index throughout the year account for a fraction $a$ of its market value, while deletions account for a fraction $d$. The difference in the market values of additions and deletions $(a - d)$ is due to mergers between index firms. There is an index fund that mechanically tracks the index, holding all the index stocks at all times.

There is also an “index-neutral” fund that mechanically buys a fraction $s$ of these large stocks, regardless of whether or not they are in the index. On the average, a fraction $s$ of these stocks would be in the index. Some fraction of these stocks disappear due to mergers, some randomly selected stocks are sold for other reasons, and some randomly selected new stocks are added to the portfolio. The fund wants to keep the fraction of index stocks $s$ constant in its
portfolio.

The index fund starts by investing all of its wealth in the index. During the year, stocks worth a fraction \( d \) are deleted from the index and they have to be sold. As they are deleted from the index, they also lose the index premium \( p \), so the index fund will only get a price \( \frac{d}{1+p} \) for the stocks that it bought at a price \( d \). To replace the deleted and merged stocks, new stocks worth \( a \) are added to the index, and they are bought by the index fund at a price \( a \) (which includes the index premium – before the addition these stocks were worth \( \frac{a}{1+p} \)). We assume a merger does not on average change the combined value of the merged firms, so mergers between index firms will not affect the value of the index. If stock prices on average do not move during the year and no dividends are paid, the net portfolio value will change from 1 to

\[
1 - d + \frac{d}{1+p} = 1 - \frac{pd}{1+p}.
\]

This decrease in portfolio value is due to the premium \( p \) that the index fund lost when it had to sell a fraction \( d \) of its stocks without the index premium \( p \) although it had bought them at a price which did include that premium.

The other fund (the index-neutral fund) starts by having a fraction \( s \) of its wealth in index stocks and a fraction \( 1-s \) in non-index stocks. As the fund tries to keep its index share constant, it effectively has two portfolios: one purely consisting of index stocks and one purely consisting of non-index stocks. The index stock portfolio behaves just as the portfolio of the index fund: after a year, its value will change from 1 to \( 1 - \frac{pd}{1+p} \). The value of the non-index portfolio is not affected by the fund buying or selling non-index stocks, but it is affected by additions of non-index stocks to the index. Index additions are worth a fraction \( a \) of the index, and since the index stocks are a fraction \( s \) and the non-index stocks are a fraction \( 1-s \) of the pool of large stocks, a fraction worth \( \frac{s}{1-s}a \) of non-index stocks have to be selected. Before index addition, these stocks did not have the index premium, so they were worth \( \frac{s}{1-s} \frac{a}{1+p} \) of the non-index stocks. Hence, this fraction of the non-index portfolio will suddenly earn the premium \( p \), increasing the non-index portfolio value from 1 to

\[
1 + \frac{s}{1-s} \frac{pa}{1+p}.
\]

The total value of the portfolio of the index-neutral fund is then

\[
s \left( 1 - \frac{pd}{1+p} \right) + (1-s) \left( 1 + \frac{s}{1-s} \frac{pa}{1+p} \right) = 1 + \frac{ps(a-d)}{1+p}.
\]

In other words, the fund gains the index premium on index additions but loses it on index
deletions, so the net effect is that the fund gains the index premium on the difference between index additions and deletions (i.e. the merged stocks).

The difference in the performances of the index fund and the index-neutral fund is given by the difference in net portfolio value. This is

\[
(1 + \frac{ps(a - d)}{1 + p}) - (1 - \frac{pd}{1 + p}) = \frac{p[s(a - d) + d]}{1 + p}.
\]  

(1.2)

This expression gives us the cost of mechanically tracking the index as opposed to following a strategy where we buy similar stocks but do not make any effort to distinguish between index stocks and non-index stocks. The two portfolios will be essentially the same, except that the indexing strategy persistently loses the index premium on stocks deleted from the index while the index-neutral strategy also buys some non-index stocks and hence gains the index premium on the stocks later added to the index.\(^{10,11}\)

Note that the above example assumes a permanent index premium. If the index effects are fully reversed, the index fund would have to pay a separate cost for both additions and deletions, which would increase it to \(p\left(\frac{a + d}{1 + p}\right)\), i.e. the premium times the “non-index value” of the additions and deletions.

**Magnitude**

Table 1.13 shows the market share of the index accounted for by additions and deletions. There are fewer deletions than additions due to mergers between index firms, but their much smaller

\(^{10}\)This example ignores any portfolio returns from cash dividends or share repurchases. If two identical firms pay the same cash dividend but one of them is trading at a price premium \(p\), clearly the part of stock return that is due to dividends is diminished. Using a recent S&P 500 dividend yield figure of 1.3% and a 14% index premium, we get a 16 basis point difference in annual return. Using a recent S&P 500 P/E ratio of 26.66 and a historical payout ratio of 0.51 from the last 50 years, we get a 23 basis point difference. But since our alternative index-neutral strategy would also hold a fraction \(s\) of the portfolio in index stocks, the difference between the annual returns on an index portfolio and an index-neutral portfolio would be less than the above estimates. The price effect for cash dividends would be there and it would hurt the index fund, but it would be smaller than the direct index turnover effects discussed above, and for simplicity we ignore it in the calculations above.

\(^{11}\)The example also makes the assumption that the market value of the index stocks and the amount of money invested in index funds grow at roughly the same rate, keeping the relative index premium \(p\) more or less constant. In this case the rate was \(a - d\), the difference between the market values of index additions and deletions. If instead the index funds grow faster than the index, this would produce more one-time capital gains for the index stocks, while a slower growth rate of index funds would lead to a diminishing index premium and capital losses on index stocks.
turnover shows that individual deletions are also smaller than deletions. Hence, while market value is a very important characteristic for additions, the S&P seems to be reluctant to delete firms from the index based on their market value unless it really falls to a low level.

In 2000, there were 50 pure additions to the index, accounting for 5.6% of its market cap, and 23 pure deletions, accounting for 0.4% of its market cap. The S&P 500 index covers about 76% of the total value of the U.S. stock market. Since the index is selected primarily from the largest 1,000 firms (94% of index firms and 99.8% of index cap), we could restrict ourselves to this subset of the market. This would change the index share $s$ from 76% to about 83%.

Using $p = 16.3\%$ (the weighted average of the addition and deletion premium), $a = 5.6\%$, $d = 0.46\%$, and $s = 83\%$ and plugging these numbers into the formula, we get an annual index turnover cost of 0.66%. An additional 3-5 basis points could be added to this estimate due to cash dividends and share repurchases, bringing the total cost to around 0.7%. If all of the index premium were temporary, the cost would further increase to 0.85% per year. Naturally the cost would be lower for years with less turnover in the index.

If the share of stock market wealth indexed to the S&P 500 keeps growing, the index premium will most likely continue to grow as well, and this will further increase the index turnover cost in the future.

1.5.2 Russell 2000

Source of the Cost

Since the Russell 2000 index is selected differently from the S&P 500, we need to determine the cost somewhat differently as well.

An index fund will always hold the index. To maintain the index portfolio, once a year it has to buy additions and sell deletions. Denote the “index value” of deletions by $d$ (from below) and $d_t$ (from above), and additions by $a$ (from below) and $a_t$ (from above) This generates a cash flow of

$$d (1 - p) + d_t - a - a_t,$$

when the deletions from below lose the index premium $p$ and when the additions and deletions from above do not experience any price effects.
Now assume a "home-made index fund" that follows the rule for Russell 2000 but instead updates its own portfolio on Dec 31, i.e. 6 months off the official cycle. These two portfolios will be essentially indistinguishable from one another, except for the phase shift in the updating of the portfolios. This would be the appropriate index-neutral benchmark for the Russell 2000.

Further assume for simplicity that there is a steady stream of would-be additions to and deletions from the index, so that in 6 months we accumulate about half of the index changes that we would accumulate over 12 months. Hence, the cash flow generated by holding the index-neutral portfolio would be

\[ \frac{1}{2} [d(1-p) + d_t - a - a_t] + \frac{1}{2} [d + d_t - a(1-p) - a_t]. \]

In other words, the index-neutral portfolio would update half of the stocks after the last official change (and get the same cash flow as the index fund), and half of the stocks ahead of the next official change (selling deletions with the premium and buying additions without it). The difference between the two would be

\[ \frac{p}{2} (a + d), \]

i.e. the index-neutral investor would only suffer the premium for half of the additions and deletions. If the price effects reversed fully over 6 months, the difference would increase to \( p(a + d) \).

**Magnitude**

Turnover for the Russell 2000 is much higher than that for the S&P 500 (Table 1.14), with additions and deletions accounting for 19.9% and 4.8% from below, and 12.6% and 26.7% from above. The weighted average of the price premium for additions and deletions in 2000 is \( p = 8.9\% \). This produces an annual index turnover cost of 1.10% for the index fund relative to the similar index-neutral strategy.\(^{12}\) If the price impact reverses fully, the turnover cost increases to 2.11%.

\(^{12}\)Plus a few basis points for the difference due to dividends.
1.6 Conclusions

In this chapter we document a growing index premium for the S&P 500 and Russell 2000. The growing premium seems to be related to the growing popularity of index funds over the 1980s and 1990s. Our estimated price elasticity of demand for S&P 500 stocks, $-0.7$ using data from 2000, is close to Shleifer's (1986) estimate of unit elasticity for the same index at a time when indexing was much less popular. The elasticity of demand for Russell 2000 stocks seems to be about one third of this, or $-0.2$, so in this sense the market seems to be more efficient for S&P 500 stocks.

The price premia for additions to the two indices are 15.1% and 9.5% in 2000, with similar figures for deletions. Surprisingly though, the premia seem to come about only gradually between the announcement and effective days, as opposed to being fully incorporated into prices right after announcement. This is particularly unexpected as both researchers and practitioners have been aware of some of these S&P 500 price effects for a long time.

The cross-sectional link between the index premium (or equivalently, the slope of the demand curve) and idiosyncratic risk is clearly distinguishable for both datasets. It is significant both statistically and economically, translating to a 2% additional price premium for an S&P 500 addition with a 10% increase in annual idiosyncratic volatility. The link between the index premium and market equity also seems to be there, implying that smaller stocks would have steeper demand curves.

The annual index turnover cost, using the most recent year in the data, is about 70 bp for the S&P 500 and 110 bp for the Russell 2000. If the index premium fully reverses over the next few months, the cost increases to about 85 bp and 221 bp, respectively. This certainly seems like something investors should care about, since it acts as a guaranteed drag on returns in a way similar to the annual management fee, which in turn is only 18 bp for presumably cost-conscious Vanguard 500 investors.
1.7 References


Goetzmann, W.N. and M. Massa, 1999, “Index Funds and Stock Market Growth,” working paper, Yale School of Management and INSEAD.


## 1.8 Tables

<table>
<thead>
<tr>
<th>Year</th>
<th>N</th>
<th>CAR</th>
<th>stderr</th>
<th>N</th>
<th>CAR</th>
<th>stderr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>11</td>
<td>0.002</td>
<td>0.024</td>
<td>7</td>
<td>-0.277</td>
<td>0.102</td>
</tr>
<tr>
<td>1991</td>
<td>10</td>
<td>0.129</td>
<td>0.024</td>
<td>5</td>
<td>-0.469</td>
<td>0.152</td>
</tr>
<tr>
<td>1992</td>
<td>6</td>
<td>-0.026</td>
<td>0.040</td>
<td>5</td>
<td>-0.0409</td>
<td>0.090</td>
</tr>
<tr>
<td>1993</td>
<td>9</td>
<td>0.061</td>
<td>0.034</td>
<td>6</td>
<td>-0.053</td>
<td>0.046</td>
</tr>
<tr>
<td>1994</td>
<td>15</td>
<td>0.063</td>
<td>0.021</td>
<td>13</td>
<td>0.031</td>
<td>0.043</td>
</tr>
<tr>
<td>1995</td>
<td>23</td>
<td>0.079</td>
<td>0.018</td>
<td>14</td>
<td>-0.096</td>
<td>0.035</td>
</tr>
<tr>
<td>1996</td>
<td>21</td>
<td>0.094</td>
<td>0.024</td>
<td>14</td>
<td>-0.059</td>
<td>0.025</td>
</tr>
<tr>
<td>1997</td>
<td>24</td>
<td>0.110</td>
<td>0.018</td>
<td>7</td>
<td>-0.068</td>
<td>0.048</td>
</tr>
<tr>
<td>1998</td>
<td>37</td>
<td>0.135</td>
<td>0.017</td>
<td>10</td>
<td>-0.138</td>
<td>0.046</td>
</tr>
<tr>
<td>1999</td>
<td>39</td>
<td>0.105</td>
<td>0.022</td>
<td>11</td>
<td>-0.155</td>
<td>0.049</td>
</tr>
<tr>
<td>2000</td>
<td>50</td>
<td>0.151</td>
<td>0.029</td>
<td>23</td>
<td>-0.189</td>
<td>0.041</td>
</tr>
</tbody>
</table>

Table 1.1: Abnormal returns for S&P 500 additions and deletions. The table shows the buy-and-hold abnormal returns for all qualifying event stocks from 15 trading days before the announcement day to the effective day of the change. Stocks are given equal portfolio weights at the beginning. Stocks need to exist in CRSP 15 trading days before the announcement and 15 trading days after the index change. The abnormal return on a stock is the difference between the stock return and the return on the S&P 500 index.
| Year | Additions |          |          |  | Deletions |          |          |
|------|-----------|----------|----------| |           |----------|----------|
|      | N          | CAR      | stderr   |  | N          | CAR      | stderr   |
| 1980 | 12         | 0.047    | 0.022    | 0 | 0.000      | 0.000    |
| 1981 | 19         | 0.057    | 0.020    | 2 | -0.028     | 0.056    |
| 1982 | 26         | 0.045    | 0.017    | 6 | 0.022      | 0.051    |
| 1983 | 10         | 0.015    | 0.029    | 6 | -0.050     | 0.021    |
| 1984 | 30         | 0.024    | 0.014    | 4 | 0.071      | 0.100    |
| 1985 | 27         | 0.023    | 0.014    | 5 | -0.231     | 0.050    |
| 1986 | 28         | 0.033    | 0.014    | 8 | -0.060     | 0.043    |
| 1987 | 25         | 0.046    | 0.016    | 5 | -0.103     | 0.061    |
| 1988 | 24         | 0.040    | 0.016    | 15| -0.068     | 0.052    |
| 1989 | 28         | 0.045    | 0.012    | 16| -0.061     | 0.033    |
| 1990 | 11         | -0.011   | 0.025    | 7 | -0.260     | 0.105    |
| 1991 | 10         | 0.107    | 0.025    | 5 | -0.367     | 0.157    |
| 1992 | 6          | -0.019   | 0.042    | 5 | -0.344     | 0.092    |
| 1993 | 9          | 0.049    | 0.035    | 6 | -0.054     | 0.047    |
| 1994 | 15         | 0.059    | 0.022    | 13| 0.048      | 0.044    |
| 1995 | 23         | 0.072    | 0.018    | 14| -0.085     | 0.036    |
| 1996 | 21         | 0.081    | 0.025    | 14| -0.044     | 0.026    |
| 1997 | 24         | 0.105    | 0.019    | 7 | -0.063     | 0.050    |
| 1998 | 37         | 0.119    | 0.018    | 10| -0.113     | 0.047    |
| 1999 | 39         | 0.098    | 0.023    | 11| -0.153     | 0.050    |
| 2000 | 50         | 0.133    | 0.029    | 23| -0.198     | 0.042    |

Table 1.2: Abnormal returns for S&P 500 additions and deletions. The table shows the buy-and-hold abnormal returns for all qualifying event stocks from 15 trading days before the announcement day to 1 trading day after the effective day. (Before October 1989, changes became effective immediately as they were announced after the close of trading.) Stocks are given equal portfolio weights at the beginning. Stocks need to exist in CRSP 15 trading days before the announcement and 15 trading days after the index change. The abnormal return on a stock is the difference between the stock return and the return on the S&P 500 index.
<table>
<thead>
<tr>
<th>Year</th>
<th>Additions</th>
<th></th>
<th></th>
<th>Deletions</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>CAR</td>
<td>stderr</td>
<td>N</td>
<td>CAR</td>
<td>stderr</td>
</tr>
<tr>
<td>1987</td>
<td>224</td>
<td>0.030</td>
<td>0.017</td>
<td>241</td>
<td>-0.039</td>
<td>0.014</td>
</tr>
<tr>
<td>1988</td>
<td>171</td>
<td>0.056</td>
<td>0.015</td>
<td>170</td>
<td>-0.071</td>
<td>0.017</td>
</tr>
<tr>
<td>1989</td>
<td>100</td>
<td>0.005</td>
<td>0.020</td>
<td>82</td>
<td>-0.043</td>
<td>0.031</td>
</tr>
<tr>
<td>1990</td>
<td>283</td>
<td>0.026</td>
<td>0.012</td>
<td>154</td>
<td>-0.032</td>
<td>0.054</td>
</tr>
<tr>
<td>1991</td>
<td>377</td>
<td>0.023</td>
<td>0.010</td>
<td>330</td>
<td>-0.060</td>
<td>0.025</td>
</tr>
<tr>
<td>1992</td>
<td>408</td>
<td>0.021</td>
<td>0.011</td>
<td>381</td>
<td>-0.039</td>
<td>0.018</td>
</tr>
<tr>
<td>1993</td>
<td>318</td>
<td>0.016</td>
<td>0.010</td>
<td>308</td>
<td>-0.039</td>
<td>0.012</td>
</tr>
<tr>
<td>1994</td>
<td>399</td>
<td>0.004</td>
<td>0.011</td>
<td>401</td>
<td>-0.029</td>
<td>0.011</td>
</tr>
<tr>
<td>1995</td>
<td>322</td>
<td>0.041</td>
<td>0.012</td>
<td>238</td>
<td>-0.067</td>
<td>0.017</td>
</tr>
<tr>
<td>1996</td>
<td>364</td>
<td>0.029</td>
<td>0.012</td>
<td>304</td>
<td>-0.060</td>
<td>0.012</td>
</tr>
<tr>
<td>1997</td>
<td>385</td>
<td>0.057</td>
<td>0.010</td>
<td>305</td>
<td>-0.059</td>
<td>0.018</td>
</tr>
<tr>
<td>1998</td>
<td>368</td>
<td>0.040</td>
<td>0.011</td>
<td>223</td>
<td>0.015</td>
<td>0.017</td>
</tr>
<tr>
<td>1999</td>
<td>322</td>
<td>0.089</td>
<td>0.038</td>
<td>226</td>
<td>-0.075</td>
<td>0.031</td>
</tr>
<tr>
<td>2000</td>
<td>442</td>
<td>0.095</td>
<td>0.035</td>
<td>318</td>
<td>-0.080</td>
<td>0.024</td>
</tr>
</tbody>
</table>

Table 1.3: Abnormal returns for Russell 2000 additions and deletions that cross the lower cutoff of the index. The table shows the buy-and-hold abnormal returns for all qualifying event stocks from May 31 (when the new index is determined) to June 30 (when the new index becomes effective). Stocks are given equal portfolio weights on May 31. Stocks need to have at least 2 months of CRSP return data prior to the event. The abnormal return on a stock is the difference between the stock return and the return on its control portfolio which has similar idiosyncratic risk and market equity.
| Year | Additions | | | | | Deletions | | | |
|---|---|---|---|---|---|---|---|---|
|   | N  | CAR | stderr | N  | CAR | stderr |
| 1987 | 61 | -0.014 | 0.017 | 55 | 0.068 | 0.022 |
| 1988 | 46 | -0.032 | 0.017 | 48 | 0.031 | 0.017 |
| 1989 | 37 | 0.069 | 0.017 | 41 | -0.041 | 0.017 |
| 1990 | 59 | 0.029 | 0.019 | 73 | -0.016 | 0.013 |
| 1991 | 78 | 0.025 | 0.019 | 67 | -0.025 | 0.023 |
| 1992 | 80 | 0.014 | 0.025 | 56 | -0.032 | 0.019 |
| 1993 | 92 | -0.036 | 0.015 | 60 | 0.007 | 0.016 |
| 1994 | 95 | 0.007 | 0.016 | 65 | -0.021 | 0.020 |
| 1995 | 67 | 0.008 | 0.016 | 78 | 0.011 | 0.015 |
| 1996 | 104 | 0.003 | 0.011 | 91 | -0.016 | 0.016 |
| 1997 | 93 | 0.012 | 0.026 | 100 | -0.020 | 0.012 |
| 1998 | 91 | -0.047 | 0.014 | 108 | 0.032 | 0.011 |
| 1999 | 94 | 0.003 | 0.036 | 123 | -0.027 | 0.044 |
| 2000 | 136 | -0.040 | 0.034 | 117 | 0.014 | 0.051 |

Table 1.4: Abnormal returns for Russell 2000 additions and deletions that cross the upper cutoff of the index. The table shows the buy-and-hold abnormal returns for all qualifying event stocks from May 31 (when the new index is determined) to June 30 (when the new index becomes effective). Stocks are given equal portfolio weights on May 31. Stocks need to have at least 2 months of CRSP return data prior to the event. The abnormal return on a stock is the difference between the stock return and the return on its control portfolio which has similar idiosyncratic risk and market equity.
<table>
<thead>
<tr>
<th>Size</th>
<th>Idiosyncratic risk</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>High</td>
</tr>
<tr>
<td>Small</td>
<td>655</td>
<td>654</td>
<td>655</td>
<td>654</td>
<td>654</td>
</tr>
<tr>
<td>2</td>
<td>172</td>
<td>172</td>
<td>172</td>
<td>172</td>
<td>171</td>
</tr>
<tr>
<td>3</td>
<td>108</td>
<td>108</td>
<td>108</td>
<td>108</td>
<td>108</td>
</tr>
<tr>
<td>4</td>
<td>76</td>
<td>76</td>
<td>76</td>
<td>76</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>59</td>
<td>59</td>
<td>59</td>
<td>59</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>7</td>
<td>42</td>
<td>41</td>
<td>42</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>8</td>
<td>39</td>
<td>39</td>
<td>39</td>
<td>39</td>
<td>38</td>
</tr>
<tr>
<td>9</td>
<td>35</td>
<td>35</td>
<td>34</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>Big</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>33</td>
</tr>
</tbody>
</table>

Table 1.5: Median number of stocks in each benchmark portfolio. The statistics are computed for all U.S. firms listed on the NYSE, AMEX, or Nasdaq, and over the period 1987-2000. The 10×5 benchmark portfolios for month $t$ are formed first by dividing stocks into size deciles based on Fama-French breakpoints for market equity at the end of month $t - 2$. Stocks within each size decile are then sorted into quintiles based on the root mean squared error of a regression of stock returns on the three factors of Fama-French over the 6-month-period from month $t - 7$ to month $t - 2$. 
<table>
<thead>
<tr>
<th>Size</th>
<th>Low</th>
<th>3</th>
<th>4</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>35</td>
<td>29</td>
<td>20</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>109</td>
<td>102</td>
<td>101</td>
<td>98</td>
</tr>
<tr>
<td>3</td>
<td>202</td>
<td>198</td>
<td>195</td>
<td>191</td>
</tr>
<tr>
<td>4</td>
<td>325</td>
<td>324</td>
<td>319</td>
<td>319</td>
</tr>
<tr>
<td>5</td>
<td>516</td>
<td>513</td>
<td>507</td>
<td>510</td>
</tr>
<tr>
<td>6</td>
<td>805</td>
<td>799</td>
<td>783</td>
<td>776</td>
</tr>
<tr>
<td>7</td>
<td>1,223</td>
<td>1,228</td>
<td>1,208</td>
<td>1,229</td>
</tr>
<tr>
<td>8</td>
<td>2,089</td>
<td>2,017</td>
<td>2,026</td>
<td>1,951</td>
</tr>
<tr>
<td>9</td>
<td>3,695</td>
<td>3,931</td>
<td>3,726</td>
<td>3,593</td>
</tr>
<tr>
<td>Big</td>
<td>18,541</td>
<td>14,558</td>
<td>12,478</td>
<td>11,712</td>
</tr>
</tbody>
</table>

Table 1.6: Median market capitalization of each benchmark portfolio. The statistics are computed for all U.S. firms listed on the NYSE, AMEX, or Nasdaq, and over the period 1987-2000. The $10 \times 5$ benchmark portfolios for month $t$ are formed first by dividing stocks into size deciles based on Fama-French breakpoints for market equity at the end of month $t-2$. Stocks within each size decile are then sorted into quintiles based on the root mean squared error of a regression of stock returns on the three factors of Fama-French over the 6-month-period from month $t-7$ to month $t-2$. 
<table>
<thead>
<tr>
<th>Size</th>
<th>Low</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>0.32</td>
<td>0.54</td>
<td>0.73</td>
<td>0.96</td>
<td>1.55</td>
</tr>
<tr>
<td>2</td>
<td>0.27</td>
<td>0.39</td>
<td>0.49</td>
<td>0.61</td>
<td>0.85</td>
</tr>
<tr>
<td>3</td>
<td>0.23</td>
<td>0.33</td>
<td>0.43</td>
<td>0.53</td>
<td>0.74</td>
</tr>
<tr>
<td>4</td>
<td>0.21</td>
<td>0.30</td>
<td>0.37</td>
<td>0.47</td>
<td>0.66</td>
</tr>
<tr>
<td>5</td>
<td>0.19</td>
<td>0.27</td>
<td>0.35</td>
<td>0.43</td>
<td>0.59</td>
</tr>
<tr>
<td>6</td>
<td>0.17</td>
<td>0.25</td>
<td>0.31</td>
<td>0.38</td>
<td>0.54</td>
</tr>
<tr>
<td>7</td>
<td>0.16</td>
<td>0.23</td>
<td>0.28</td>
<td>0.35</td>
<td>0.48</td>
</tr>
<tr>
<td>8</td>
<td>0.16</td>
<td>0.21</td>
<td>0.26</td>
<td>0.32</td>
<td>0.45</td>
</tr>
<tr>
<td>9</td>
<td>0.15</td>
<td>0.20</td>
<td>0.23</td>
<td>0.28</td>
<td>0.39</td>
</tr>
<tr>
<td>Big</td>
<td>0.15</td>
<td>0.18</td>
<td>0.21</td>
<td>0.25</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Table 1.7: Median idiosyncratic risk for each benchmark portfolio. Idiosyncratic risk is defined as the root mean squared error of a regression of stock returns on the three factors of Fama-French. The statistics are computed for all U.S. firms listed on the NYSE, AMEX, or Nasdaq, and over the period 1987-2000. The $10 \times 5$ benchmark portfolios for month $t$ are formed first by dividing stocks into size deciles based on Fama-French breakpoints for market equity at the end of month $t-2$. Stocks within each size decile are then sorted into quintiles based on idiosyncratic risk computed over the 6-month-period from month $t-7$ to month $t-2$. 
### Share of stocks with zero trading volume

<table>
<thead>
<tr>
<th>Size</th>
<th>Low</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>20.2%</td>
<td>11.2%</td>
<td>8.3%</td>
<td>9.1%</td>
<td>13.2%</td>
</tr>
<tr>
<td>2</td>
<td>8.7%</td>
<td>5.3%</td>
<td>4.3%</td>
<td>3.5%</td>
<td>3.3%</td>
</tr>
<tr>
<td>3</td>
<td>3.9%</td>
<td>2.9%</td>
<td>2.3%</td>
<td>1.4%</td>
<td>1.2%</td>
</tr>
<tr>
<td>4</td>
<td>2.2%</td>
<td>1.3%</td>
<td>1.1%</td>
<td>1.0%</td>
<td>1.1%</td>
</tr>
<tr>
<td>5</td>
<td>0.7%</td>
<td>0.5%</td>
<td>0.4%</td>
<td>0.2%</td>
<td>0.4%</td>
</tr>
<tr>
<td>6</td>
<td>0.5%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>0.1%</td>
</tr>
<tr>
<td>7</td>
<td>0.4%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>8</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>9</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Big</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Table 1.8: Median share of stock-days with zero trading volume in each benchmark portfolio. The statistics are computed for all U.S. firms listed on the NYSE, AMEX, or Nasdaq, and over the period 1987-2000. The $10 \times 5$ benchmark portfolios for month $t$ are formed first by dividing stocks into size deciles based on Fama-French breakpoints for market equity at the end of month $t - 2$. Stocks within each size decile are then sorted into quintiles based on the root mean squared error of a regression of stock returns on the three factors of Fama-French over the 6-month-period from month $t - 7$ to month $t - 2$.

### Number of stocks

<table>
<thead>
<tr>
<th>Size</th>
<th>Low</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>10.1</td>
<td>15.4</td>
<td>14.1</td>
<td>12.2</td>
<td>6.1</td>
</tr>
<tr>
<td>2</td>
<td>17.5</td>
<td>19.8</td>
<td>25.1</td>
<td>25.6</td>
<td>25.3</td>
</tr>
<tr>
<td>3</td>
<td>10.1</td>
<td>11.6</td>
<td>16.7</td>
<td>21.8</td>
<td>28.1</td>
</tr>
<tr>
<td>4</td>
<td>4.4</td>
<td>3.6</td>
<td>6.5</td>
<td>9.6</td>
<td>17.4</td>
</tr>
<tr>
<td>5</td>
<td>1.3</td>
<td>2.6</td>
<td>3.0</td>
<td>6.2</td>
<td>10.4</td>
</tr>
<tr>
<td>6</td>
<td>0.4</td>
<td>0.3</td>
<td>0.4</td>
<td>1.4</td>
<td>4.5</td>
</tr>
<tr>
<td>7</td>
<td>0.1</td>
<td>0.1</td>
<td>1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.9: Russell 2000 additions in 1987-2000. This is the average number of addition event stocks per year for each benchmark portfolio. The benchmark portfolios are defined as before.
Table 1.10: The results of a Fama-MacBeth regression for 14 annual cross-sections from 1987 to 2000. For each cross-section, we regress the cumulative abnormal returns on individual stocks in June on the stocks' market capitalizations and idiosyncratic volatilities estimated from November 1 to April 30. We then compute the time-series means of regression coefficients $\gamma_{0t}$, $\gamma_{1t}$, and $\gamma_{2t}$. Standard errors for the means are reported in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>$\bar{\gamma}_{0t}$</th>
<th>$\bar{\gamma}_{1t}$</th>
<th>$\bar{\gamma}_{2t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.016</td>
<td>0.039</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.009)</td>
<td>(0.011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.081</td>
<td>-0.027</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.033)</td>
<td>(0.015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.054</td>
<td>0.032</td>
<td>-0.023</td>
<td></td>
</tr>
<tr>
<td>(0.041)</td>
<td>(0.013)</td>
<td>(0.016)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.11: S&P 500 additions in 1990-2000. This is the number of addition event stocks for each benchmark portfolio. The benchmark portfolios are defined as before.

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>5</td>
<td>11</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>12</td>
<td>23</td>
<td>42</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>13</td>
<td>35</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

41
Table 1.12: The results of cumulative abnormal returns for index additions and deletions in 1990-2000 regressed on the log of market equity and idiosyncratic volatility of a stock as well as year dummies. The cumulative abnormal returns are defined as the market-adjusted cumulative returns from 15 trading days before the announcement up to the effective day of the change. White’s heteroskedasticity-consistent standard errors are reported in parentheses.

<table>
<thead>
<tr>
<th>Year</th>
<th>Additions</th>
<th>Deletions</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>turnover</td>
<td>N</td>
</tr>
<tr>
<td>1980</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>1981</td>
<td>19</td>
<td>0.026</td>
</tr>
<tr>
<td>1982</td>
<td>26</td>
<td>0.019</td>
</tr>
<tr>
<td>1983</td>
<td>10</td>
<td>0.010</td>
</tr>
<tr>
<td>1984</td>
<td>30</td>
<td>0.026</td>
</tr>
<tr>
<td>1985</td>
<td>27</td>
<td>0.021</td>
</tr>
<tr>
<td>1986</td>
<td>28</td>
<td>0.020</td>
</tr>
<tr>
<td>1987</td>
<td>24</td>
<td>0.017</td>
</tr>
<tr>
<td>1988</td>
<td>24</td>
<td>0.024</td>
</tr>
<tr>
<td>1989</td>
<td>28</td>
<td>0.020</td>
</tr>
<tr>
<td>1990</td>
<td>11</td>
<td>0.006</td>
</tr>
<tr>
<td>1991</td>
<td>10</td>
<td>0.018</td>
</tr>
<tr>
<td>1992</td>
<td>6</td>
<td>0.005</td>
</tr>
<tr>
<td>1993</td>
<td>9</td>
<td>0.017</td>
</tr>
<tr>
<td>1994</td>
<td>15</td>
<td>0.032</td>
</tr>
<tr>
<td>1995</td>
<td>23</td>
<td>0.028</td>
</tr>
<tr>
<td>1996</td>
<td>21</td>
<td>0.029</td>
</tr>
<tr>
<td>1997</td>
<td>24</td>
<td>0.021</td>
</tr>
<tr>
<td>1998</td>
<td>37</td>
<td>0.048</td>
</tr>
<tr>
<td>1999</td>
<td>39</td>
<td>0.032</td>
</tr>
<tr>
<td>2000</td>
<td>50</td>
<td>0.056</td>
</tr>
</tbody>
</table>

Table 1.13: Index turnover for S&P 500 additions and deletions. The table shows the percentage share of additions to and deletions from the index as they occur at the close of trading on effective day. This table only includes stocks that exist both before and after the event, which eliminates virtually all changes due to mergers and acquisitions.
<table>
<thead>
<tr>
<th>Year</th>
<th>Additions From below</th>
<th>From above</th>
<th>Deletions From below</th>
<th>From above</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N weight</td>
<td></td>
<td>N weight</td>
<td></td>
</tr>
<tr>
<td>1987</td>
<td>246 8.4 61 7.9</td>
<td></td>
<td>252 3.3 55 10.2</td>
<td></td>
</tr>
<tr>
<td>1988</td>
<td>189 7.5 46 6.1</td>
<td></td>
<td>185 3.0 49 8.6</td>
<td></td>
</tr>
<tr>
<td>1989</td>
<td>103 1.7 37 5.6</td>
<td></td>
<td>100 1.5 41 7.7</td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>284 7.3 61 8.2</td>
<td></td>
<td>173 2.3 74 15.5</td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td>403 18.3 78 8.8</td>
<td></td>
<td>344 2.3 67 16.8</td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td>439 17.4 81 10.2</td>
<td></td>
<td>388 3.8 56 12.3</td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td>343 13.0 93 10.6</td>
<td></td>
<td>314 3.9 60 13.0</td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td>461 18.0 97 10.0</td>
<td></td>
<td>414 6.4 65 13.3</td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>363 11.6 67 6.9</td>
<td></td>
<td>251 3.9 79 13.5</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>413 14.1 106 10.9</td>
<td></td>
<td>312 4.8 91 16.4</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>435 14.3 94 7.3</td>
<td></td>
<td>320 4.7 101 15.8</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>425 13.8 97 8.9</td>
<td></td>
<td>243 4.2 108 17.7</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>410 13.7 101 10.2</td>
<td></td>
<td>248 3.9 124 23.8</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>543 19.9 137 12.6</td>
<td></td>
<td>334 4.8 117 26.7</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.14: Index turnover for Russell 2000 additions and deletions. The table shows the percentage share of additions to the index as they occur (June 30), and the share of deletions the month before they occur (May 31). In a typical year we miss about 5 stocks because we cannot find them in CRSP.
1.9 Figures

Figure 1-1: S&P 500 index additions and deletions in 2000. The figure represents market-adjusted cumulative abnormal returns for buy-and-hold portfolios in event time. Index changes are announced after close on day $-5$, and they become effective after close on day $0$. Since the difference between the two days may be longer or shorter for some stocks, we have shrunk or stretched the actual intermediate time to fit the 5-day interval.
Figure 1-2: S&P 500 index additions and deletions in 1999. The figure represents market-adjusted cumulative abnormal returns for buy-and-hold portfolios in event time. Index changes are announced after close on day -5, and they become effective after close on day 0. Since the difference between the two days may be longer or shorter for some stocks, we have shrunk or stretched the actual intermediate time to fit the 5-day interval.
Figure 1-3: S&P 500 index additions and deletions in 2000. The figure represents long-horizon market-adjusted cumulative abnormal returns for buy-and-hold portfolios in event time. Index changes become effective after close on day 0.
Figure 1-4: S&P 500 index additions and deletions in 1999. The figure represents long-horizon market-adjusted cumulative abnormal returns for buy-and-hold portfolios in event time. Index changes become effective after close on day 0.
Figure 1-5: S&P 500 index premium for additions in 1980-2000. This figure shows the CARs of additions from 20 trading days before the change to market close 1 trading day after the change. Prior to October 1989, there was no advance notice of index changes. Each CAR represents the buy-and-hold portfolio return in event time for qualifying additions, equal-weighted at trading day −20.
Figure 1-6: Russell 2000 index additions and deletions in 1987-2000. The figure represents the cumulative abnormal returns in June on buy-and-hold portfolios of event stocks. Only the additions and deletions crossing the lower index cutoff are shown.
Figure 1-7: Russell 2000 index additions and deletions in 2000. The figure represents the cumulative abnormal returns on buy-and-hold portfolios. Trading day -22 represents May 31, when the new index composition is determined, and trading day 0 represents June 30, when the index changes take place. Only the additions and deletions crossing the lower index cutoff are shown.
Figure 1-8: Russell 2000 index additions and deletions in 1999. The figure represents the cumulative abnormal returns on buy-and-hold portfolios. Trading day $-22$ represents May 31, when the new index composition is determined, and trading day 0 represents June 30, when the index changes take place. Only the additions and deletions crossing the lower index cutoff are shown.
Figure 1-9: Russell 2000 index additions and deletions in 2000. The figure represents the long-run cumulative abnormal returns on buy-and-hold portfolios. Trading day -22 represents May 31, when the new index composition is determined, and trading day 0 represents June 30, when the index changes take place. Only the additions and deletions crossing the lower index cutoff are shown.
Figure 1-10: Russell 2000 index additions and deletions in 1999. The figure represents the long-run cumulative abnormal returns on buy-and-hold portfolios. Trading day \(-22\) represents May 31, when the new index composition is determined, and trading day 0 represents June 30, when the index changes take place. Only the additions and deletions crossing the lower index cutoff are shown.
Chapter 2

What Makes Demand Curves for Stocks Slope Down?

2.1 Introduction

On July 9, 2002, Standard and Poor's announced that it would delete all seven non-U.S. firms from its S&P 500 index and replace them with U.S. firms. The changes were to take place after the close of trading on July 19. The deletions and additions included such large firms as Royal Dutch Petroleum, Unilever, Goldman Sachs, and UPS. The day following the announcement, the deleted firms fell by an average of 3.7% while the added firms went up by 5.9% relative to the value-weighted market index, reportedly on trading by hedge funds and active managers.\(^1\) During the ten days leading to the effective day, the cumulative market-adjusted return was \(-6.6\%\) for the deletions and \(+12.3\%\) for the additions – all on a bureaucratic event which contained absolutely no news about the level or riskiness of the cash flows of the firms involved. This event received considerable publicity as the biggest shake-up of the S&P 500 index since the break-up of AT&T in 1983, and yet it produced a very significant price impact which showed no signs of reversal at least in the following two months (Figure 2-1).

Rather than being an anomaly, this event actually illustrates the typical behavior of stocks added to or deleted from the S&P 500 index. In 2000, we observed a cumulative abnormal

\(^1\)The Wall Street Journal, 7/11/02.
return of about +15% for the 50 index additions and about -19% for the 23 deletions (Figure 1-1). Significant price effects have also been associated with a variety of other stock market indices both in the U.S. and all over the world, and a growing empirical literature has documented some of these effects.

These empirical findings have been taken as strong evidence for downward-sloping demand curves for stocks. When stocks are added to or deleted from an index, index funds mechanically tracking the index tend to buy the additions and sell the deletions as close as possible to the time of the official change in order to minimize their tracking error. For the S&P 500, mechanical indexers account for approximately 10% of the market value of every stock in the index. When demand curves for stocks slope down, the large demand shocks due to indexers can move prices and generate the observed effect.

On the other hand, the basic valuation formula of neoclassical finance tells us that price equals expected future cash flows discounted by systematic risk. The supply of a stock should not affect its price, implying that demand curves for stocks are (almost) perfectly horizontal. Any deviation from this fundamental price represents a profitable trading opportunity which would be quickly exploited and thus corrected by active market participants.

How can we reconcile this discrepancy between the predictions of neoclassical finance and the results of a growing body of empirical work?

To address this issue in the context of index changes, a variety of hypotheses have been suggested, perhaps the most prominent being liquidity, information, and market segmentation. However, it is challenging for any one of the hypotheses alone to be sufficiently general to account for the all of our empirical index evidence. Moreover, the magnitude predicted by each of them has not been explored in the literature, so it remains unclear whether any of the hypotheses can theoretically explain more than a negligible part of the index premium. Yet the puzzle about demand curves is precisely about the magnitude of the slope and not its sign.

In this chapter, our goal is to provide a theoretical explanation for downward-sloping demand curves that satisfies two important criteria: First, it is generally applicable, potentially explaining the evidence from all of the various empirical tests. Second, it can account for at least a meaningful part of the empirically observed index premium.

In a traditional CAPM benchmark, the slopes of the demand curves are determined by the
risk aversion of the representative investor. Since the representative investor is willing to hold the market portfolio, we can infer his risk aversion from the market risk premium and market volatility. With a large number of assets, the dollar variance due to market risk completely swamps the dollar variance due to the idiosyncratic risk of a single stock. This means that when the representative investor requires a certain risk premium for bearing the entire supply of market risk, he will require only a tiny risk premium for bearing (a very much smaller supply of) idiosyncratic risk. Hence, the price of a stock depends almost entirely on its systematic risk, not the supply of its idiosyncratic risk, so the demand curves for stocks should be almost perfectly horizontal and we should not observe a meaningful index premium.

To deviate from the traditional CAPM setting, we let investors invest in the stock market through financial intermediaries. Such intermediaries clearly constitute a significant part of the market: at the end of 2000, large institutions owned 55% of the market value of all stocks listed on NYSE, AMEX, and Nasdaq. While this share of institutions has grown over time, neoclassical finance never ignored it accidentally. Instead, it has generally been assumed that intermediaries would act only as a veil for the end investors, perhaps effectively making the end investors better informed but not changing their preferences, and hence they could be conveniently ignored in a model. More recently though, Allen (2001), Merton and Bodie (2002), and Shleifer and Vishny (1997) among others have focused attention on this issue, suggesting that the presence of institutions may in fact have significant implications for asset pricing. Here we build a model to explore this possibility and to see whether it has implications for demand curves for stocks.

In our model, there is a fixed cost to actively managing a stock portfolio, presumably reflecting the costs of acquiring information about the fundamentals of individual stocks, and this cost is paid only by professional money managers. End investors can invest in individual stocks only through these active managers (stock pickers) who charge a fee to cover their costs. The end investors can also invest in the market portfolio (through passive managers who charge no fee) and in the riskless asset. We do not consider agency issues, so the only real friction we introduce relative to the CAPM setting is the fixed cost and the corresponding fee for active management.

\[^2\text{Author's calculations for the CRSP universe and the Spectrum database for 13F institutions.}\]
We find that the delegation of portfolio management completely changes the cross-sectional pricing of stocks. Now the slope of the demand curve for a stock is no longer determined by end investors' risk aversion – instead it depends on the wealth allocated to active managers, which in turn depends on the fee charged by the active managers. A numerical calibration in our simple setting reveals that increasing the management fee from zero to 1.5% of the invested assets increases the slope of the demand curve roughly by a factor of 1,000, thus increasing the price impact of a demand shock by a factor of 1,000. If the fee is zero, the model collapses to the CAPM benchmark where the slope of the demand curve is determined by the risk aversion of end investors.

What is the intuition for this result? In equilibrium, the allocation of wealth to active managers is determined by their after-fees returns. End investors will have to be indifferent between investing with the active managers and investing in the passive market portfolio. The allocation to the active managers will settle at a level where the active managers earn alphas roughly equal to their management fees. This means that the demand curves for stocks will be sufficiently steep to allow for some dispersion in alphas. The equilibrium slope of the demand curve is then a measure of the equilibrium level of inefficiency in the market which allows the active managers to earn their fees. This is also consistent with the empirical results of e.g. Wermers (2000) and Daniel et al. (1997) who find that active managers outperform the market approximately by the amount of their fees.

Given the significant nonzero alphas in equilibrium, why is it that a small end investor behaves so aggressively when he has the information himself and so conservatively when he invests through a small active manager? The first-best contract between the two would involve a fixed lump-sum payment to the manager and unrestricted investment (using the manager's information) for the end investor. But since portfolios are almost costless to repackage, the absence of arbitrage enforces linear pricing, so the dollar management fee has to be approximately linear in the size of the portfolio. This takes us away from the world of first-best contracts, and it gives the end investor a reason to scale back the size of his investment in order to minimize the fee paid to the manager. Hence, the "inefficient" linearity of the management fee is what supports the equilibrium that is so different from the CAPM benchmark.

Yet it is important to realize that the institutions per se are not the source of the friction in
the model. Our model differs from that CAPM due to the fixed cost one has to pay in order to become an informed and active trader in individual stocks. When the fixed cost is large, institutions arise naturally so that all investors in the economy can share the cost through the proportional fee. In fact, if the end investors had to pay the cost themselves, demand curves in equilibrium would be even steeper than in the presence of institutions. Hence, the underlying fixed cost endogenously gives rise to institutions which actually make stock prices more efficient, i.e. closer to the CAPM benchmark.

This chapter proceeds as follows. Section 2 starts with a simple CAPM benchmark and contrasts it with the empirical evidence to illustrate the puzzle. It also briefly addresses the most prominent hypotheses in the literature to show that they do not provide easy and general answers to the puzzle. Section 3 presents our model and the equilibrium, and it provides a numerical calibration to show the magnitudes of the predicted effects. Section 4 briefly links our empirical results from Chapter 1 to the predictions made in this chapter. Section 5 discusses interpretations and extensions of the model, and section 6 concludes. The appendix presents a more elaborate (and perhaps more realistic) model to verify the robustness of the predictions from our simple model in section 3. The appendix also contains all algebra and figures.

2.2 The Puzzle: Theory and Empirical Evidence

2.2.1 Traditional Arguments

Both the CAPM and the APT tell us that the price of an average stock is equal to its expected future cash flows discounted by their systematic risk. The supply of the stock does not enter the pricing equation. In an equilibrium model such as the CAPM, the supply of the stock enters only indirectly through its effect on the pricing kernel, i.e. the marginal utility of the representative investor. When there is a large number of stocks, this indirect effect through the pricing kernel is negligible. This is why we can take the stock's beta and the market risk premium as exogenous, obtaining a pricing formula where the supply of the stock does not matter. Equivalently, we can say that the demand curve for a stock is (almost) perfectly horizontal.

None of our models of course literally implies that the demand curve for a stock is perfectly
horizontal. The real question here is about the magnitude of that slope: Is it really "negligible" as suggested by the neoclassical models, or does it deviate "significantly" from zero? In other words, can we assume for practical purposes that the stock price is unaffected by the supply of the stock? We start by presenting a simple CAPM calibration to see what exactly a negligible price impact would mean.

2.2.2 A Simple CAPM Calibration

Let there be \( N_S \) stocks with a supply of 1 unit each, and a risk-free asset with an infinitely elastic supply. One period from now each stock pays a liquidating dividend of \( \tilde{x}_i = a_i + b_i \tilde{y} + \bar{e}_i \). Systematic shocks to the economy are represented by the unexpected return on the market portfolio \( \tilde{y} \sim N(0, \sigma_m^2) \). Idiosyncratic shocks to the stock are denoted by \( \bar{e}_i \sim N(0, \sigma_i^2) \). \( a_i \) and \( b_i \) are stock-specific constants. The return on the risk-free asset is normalized to zero.

The economy is populated by mean-variance investors who can be aggregated into a representative investor with CARA utility and a coefficient of absolute risk aversion \( \gamma \).

The representative investor’s maximization problem is:

\[
\max \{ \theta_i \} \quad \mathbb{E} \left[ -\exp \left( -\gamma \tilde{W} \right) \right]
\]

s.t. \( \tilde{W} = W_0 + \sum_{i=1}^{N_S} \theta_i (\tilde{x}_i - P_i) \). \hspace{1cm} (2.1)

We calculate the first-order conditions with respect to \( \theta_i \), taking the market variance \( \sigma_m^2 \) as exogenous. We denote the equilibrium supply held by the investor as \( u_i \), and we plug it in for \( \theta_i \). This gives us the equilibrium price:

\[
P_i = a_i - \gamma \left[ \sigma_m^2 \left( \sum_{j \neq i} u_j b_j \right) b_i + \left( \sigma_m^2 b_i^2 + \sigma_i^2 \right) u_i \right]
\]

\( \text{depends on systematic risk } b_i \)

\( \text{depends on supply } u_i \) \hspace{1cm} (2.2)

\(^{3}\)Since the market return is a value-weighted return on individual stocks, the idiosyncratic stock returns actually have to add up to zero. We ignore this constraint for analytical convenience. This should have a negligible impact on our results when there is a large number of assets.
The price is equal to the expected payoff $a_i$ minus a discount, where the part of the discount that does not depend on supply contains a summation across all stocks, so the price discount will be dominated by the term that does not depend on the stock's supply.

We pick a one-year holding period, $N_S = 1,000$, $a_i = 105$, $b_i = 100$, and $\sigma_{Z_i}^2 = 900$ for all stocks and $\sigma_m^2 = 0.04$ for the market variance. We also set $\gamma = 1.247 \times 10^{-5}$ which produces an equilibrium market risk premium of 5%. Each stock will then have a price of 100, market beta of 1, and idiosyncratic standard deviation of return of 30%.

Now consider a supply shock of $-10\%$ to a stock. E.g. a new investor enters the market and buys 10% of the shares of stock $i$. Plugging in $u_i = 0.9$, the price of stock $i$ will then increase to 100.00162. In other words, this supply shock will produce a 0.16 basis point price impact. Part of this impact is due to the decreased supply of market risk and in fact all stocks would go up by 0.05 bp for this reason, so relative to the other stocks this stock would go up by 0.11 bp. This is what the "almost perfectly horizontal" demand curves mean.\footnote{These results are not affected by the choice of CARA utility as opposed to CRRA utility. See section 2.5.5 for more details on CRRA utility.}

2.2.3 Empirical Evidence

When testing for the slope of the demand curve, we cannot simply look at the market orders for a stock and compute their price impact. This would not allow us to distinguish between the price impact due to possible private information of the trader and the price impact due to the mechanical supply shock (see Kyle (1985) for an example of this setup). For this reason, most tests of demand curves focus on a subset of specific large supply shocks where the source can be identified as uninformed by both the market participants and the econometrician.

One possible sample is provided by large block trades, studied by e.g. Scholes (1972) and Holthausen and Leftwich (1987). Seasoned equity offerings provide another experiment, studied by e.g. Loderer, Cooney, and van Drumen (1991). Except for the early study by Scholes, these papers typically find relatively small negative values for the price elasticity of demand (e.g. a median of $-4.31$ and mean of $-11.1$ for Loderer et al.). Nevertheless, in these event studies it is not an easy task to control for the information conveyed by the event, and this could contribute to the relatively wide dispersion in elasticity estimates across different papers.
A cleaner approach involves changes in widely tracked stock market indices. Shleifer (1986) uses changes in the S&P 500 index and the consequent demand shocks by the investors who track the index to measure the slope of the demand curve for a stock. Several other papers have also followed this approach and documented a substantial price impact around S&P 500 index changes (e.g. Lynch and Mendenhall (1997)) which seems to have grown with the popularity of indexing (Morck and Yang (2001)). There is a growing empirical literature documenting similar effects for other indices in the U.S. (such as the Russell indices) as well as for a variety of indices around the world. The studies for the S&P 500 suggest a price elasticity of demand of approximately unity. E.g. in 2000, there was a 15% cumulative price premium for index additions while the demand shock by mechanical indexers was approximately 10% of the shares outstanding of each stock.\(^5\)

Clearly the actual estimates for the slope of the demand curve are not even remotely consistent with our simple CAPM calibration. It predicted only a 0.1 basis point price impact for a 10% demand shock, and playing with the model’s parameters will not make any meaningful changes to this enormous discrepancy. While we should not expect a perfect mapping between a simple model and reality, in this case our CAPM benchmark is obviously missing some important elements that drive the empirically observed price effect.

### 2.2.4 Suggested Hypotheses

Many different hypotheses have been suggested to explain the large slope for the demand curves for stocks in the context of index addition and deletion. Yet so far none of the papers in the literature has attempted to calibrate the commonly suggested hypotheses to actual data. Could they indeed theoretically explain more than a trivial fraction of the index premium? And how applicable are they?

It would be outside the scope of this chapter to conduct an exhaustive investigation of each hypothesis. Hence, we cannot definitively rule out any of the hypotheses, but we can point out some suggestive evidence about them.

\(^5\)The index premium was computed in Chapter 1. The size of mechanical indexers is obtained from Standard and Poor’s and the Wall Street Journal, and it matches the estimates used in other papers (e.g. Blume and Edelen (2001) and Wurgler and Zhuravskaya (2002)).
Liquidity

Stocks in the S&P 500 are typically among the most liquid stocks, perhaps due to their large size, the large pool of potential investors, and the more easily available information about them. This shows up in greater trading volume and narrower bid-ask spreads for the index stocks. Investors could rationally pay a premium for more liquid stocks because that will reduce their own adverse price impact when they sell the stock. If investors rationally anticipate this adverse price impact for themselves and for all future investors, the effect of liquidity capitalized in the stock price today might be nontrivial. If S&P 500 membership is good news for the stock’s liquidity, the price should go up even when the firm’s underlying cash flows are unaffected. Liquidity thus seems like a plausible explanation, and it could very well account for part of the observed index premium.

However, liquidity has a much harder time explaining price effects for stocks within an index, i.e. when all stocks concerned are members of the index both before and after the event. Kaul et al. (2000) investigate an event in the Toronto Stock Exchange where the public float was officially redefined, resulting in changes in index weights across index stocks. Their estimates imply a price elasticity of demand of about 0.3. Greenwood (2001) studies a large event for the Nikkei 225 index which had a very significant price impact on the stocks that were in the index before and after the event. Practitioners are well aware of intra-index price effects, e.g. as evidenced by speculative positions before Morgan Stanley Capital International redefined its indices, tracked closely by $600 billion and tracked loosely by $3 trillion, to be based on the float and not the number of shares outstanding. Hence, liquidity, as arising from index membership per se, cannot account for all these findings.

Perhaps a more complicated story could present liquidity as the main driver of these price effects, but it seems that this story would have to link liquidity to the official index weight or index fund holdings directly (after controlling for size of the firm). Yet index funds predominantly follow buy-and-hold strategies, trying to replicate the index by holding the underlying stocks as closely as possible to the official index weights, which means that greater holdings by

---

6 This is the value of $\frac{\Delta Q}{\Delta P}$ calculated by the author based on the regression estimates and a 4% market share for indexers reported in the paper.

index funds actually reduce liquidity. Thus it is not immediately obvious how to generalize the liquidity explanation to cover all these effects.

Market Segmentation

Merton (1987) suggests that the price of a stock could be increasing in its investor base. Applying his reasoning to our setup, the addition of a stock to the S&P 500 could increase its visibility to investors, make information more widely available, and allow those investors who are restricted to the S&P 500 universe to invest in the stock. Therefore it seems plausible that membership in the S&P 500 would increase the investor recognition of a stock and thus the new demand for the stock could increase its price.

While this explanation could also be part of the answer, it faces the same challenge as the liquidity hypothesis. It is easy to believe that the investor recognition of a stock depends on membership in the S&P 500, but it is much harder to explain why it would depend on the official weight within the index (controlling for size). Thus the empirically observed price effects following index weight redefinitions do not immediately follow from this reasoning.

Instead of considering shocks to the investor base, we could also look at the increased risk aversion of active investors arising from a highly segmented market. Perhaps active investors are so poorly diversified that they cannot aggressively exploit mispricings and react to uninformed supply shocks. If we try our CAPM calibration with 20 stocks instead of 1,000, we still get only a 5 basis point price impact, or if we do the calibration with 1,000 stocks, allowing each investor to know about 20 stocks and then invest passively in the remaining 980 stocks, our result is essentially unchanged, because forcing the investors to bear market risk in equilibrium makes them very aggressive in exploiting idiosyncratic mispricings. Hence, it is not clear how we could apply the market segmentation story to account for the large intra-index price effects.

Information

Perhaps addition to the S&P 500 conveys positive information about a stock. After all, it means the stock has earned the seal of approval by the S&P index selection committee. To the extent that this information was not anticipated by the market, we would expect a price increase following index addition.
S&P explicitly contradicts this by stating that index membership should not be taken as investment advice and that any anticipated stock price performance will not influence the decision on index membership. Perhaps more convincing evidence is provided by tests with other indices where index membership is based on a mechanical and transparent rule as opposed to subjective selection by a committee. For example, the Russell indices are based on a mechanical market-cap rule, and still we observe both economically and statistically significant price effects for addition to the Russell 2000 index. Practitioners also keep a close eye on changes to other mechanically determined indices such as the Nasdaq 100.\footnote{"Nasdaq 100 Index Shuffle Is Expected to Bring 13 Changes to List of Stocks," The Wall Street Journal, 11/12/2001.} Even if information does play a role in the S&P 500, it is hard to associate it with some of the other index evidence we have.

### 2.3 An Explanation with Financial Intermediaries

#### 2.3.1 Motivation

Finding the fundamental value of a firm is not an easy task. It takes time and effort to investigate the firm and its environment, including the firm's products, customers, suppliers, and competitors, and this has to be done continuously as all of these may change over time. Coming up with a meaningful valuation also requires some literacy in finance. While some individual investors are certainly capable and willing to engage in this activity, it seems plausible that most of the "smart money" in the market is invested by professionals. At the end of 2000, large institutional investors accounted for 55% of the market value of stocks traded on the NYSE, AMEX, and Nasdaq, and one could argue that these institutions represent an even greater share of relatively informed investors. Professionals at these institutions are generally the ones who are trained for the job, have easy access to information, and do the job full-time which allows them to react almost immediately to changing market prices or new information. It may be that individual investors make the market efficient not so much by trading stocks directly but by investing part of their wealth with professional active money managers.

Presumably the institutions have emerged because there is some fixed cost to becoming an
informed and active market participant. End investors then have to pay this cost as a fee for the services provided by the professional money managers. A typical actively managed U.S. equity mutual fund charges an annual fee of approximately 1.5% of assets under management.\footnote{This is perhaps the most commonly quoted value for the annual fee, but there is some dispersion here. For example, Kacperczyk, Sialm, and Zheng (2002) report that the average actively managed diversified U.S. equity fund had an expense ratio of 1.28% of assets under management in 1984-1999.} For the end investors this means they should not only consider the possible mispricing of individual stocks but also whether those mispricings are large enough to justify the costs of active management.

Could this delegation of portfolio management and the underlying fixed cost have meaningful implications for the pricing of stocks? Or will active managers simply act as a veil for end investors, giving us the same results as a representative agent setting?

\subsection{The Model}

We consider a setup (Figure 2-2) similar to the one we used earlier for the CAPM calibration. There are two differences: First, the end investors can invest in the stock market only indirectly through an active manager (a stock picker) and a passive manager (who just holds the market portfolio). This is because the end investors presumably lack the resources to select an efficient portfolio of individual stocks. Second, there are some noise traders who hold a randomly chosen portfolio of stocks. It is the deviation of these noise traders’ portfolio from the market portfolio that creates possibilities for the active managers to earn positive abnormal returns relative to the market portfolio. We abstract entirely from any potential agency issues between the money managers and the end investors.

\subsubsection{Assets}

As before, there are \( N_S \) stocks (a large number) with a supply of 1 unit each, and a risk-free asset with an infinitely elastic supply. One period from now each stock pays a liquidating dividend of \( \bar{x}_i = a_i + b_i \bar{y} + \bar{\epsilon}_i \). Systematic shocks to the economy are represented by the unexpected return on the market portfolio \( \bar{y} \sim N(0, \sigma^2_m) \). Idiosyncratic shocks to the stock are denoted by \( \bar{\epsilon}_i \sim N(0, \sigma^2_{\epsilon_i}) \). \( a_i \) and \( b_i \) are stock-specific constants. The return on the risk-free asset is normalized to zero.
When aggregating across stocks, we make two simplifying assumptions. We let all stocks have the same values of $a_i$, $b_i$, and $\sigma_{e_i}^2$. We also assume a continuum of stocks with a measure $N_S$, so that our results depend on the distribution of noise trader holdings but not on their particular realizations.

**End Investors**

The economy is populated by mean-variance investors who can be aggregated into a representative investor with CARA utility and a coefficient of absolute risk aversion $\gamma_e$. Rather than investing in individual stocks, the end investor can only pick how much to invest in an actively managed portfolio and the market portfolio, with the rest of his wealth invested in the risk-free asset. He then maximizes:

$$\max_{\{w_a,w_p\}} \mathbb{E} \left[ - \exp \left( -\gamma_e \tilde{W}_1 \right) \right]$$

s.t. $\tilde{W}_1 = W_0 + w_a \tilde{R}_a + w_p \tilde{R}_m$, \hspace{1cm} (2.3)

where $\tilde{R}_a$ and $\tilde{R}_m$ are the excess returns on the actively managed portfolio and the market portfolio, respectively, and $W_a$ and $W_p$ are the dollar allocations to each.

Denoting the excess return on stock $i$ as $\tilde{R}_i$ and the price of the market portfolio as $P_m$, we can write the portfolio returns as

$$\tilde{R}_a = \left( \sum_{i=1}^{N_S} v_i \tilde{R}_i \right) - f_a$$ \hspace{1cm} (2.4)

$$\tilde{R}_m = \frac{1}{P_m} \sum_{i=1}^{N_S} P_i \tilde{R}_i,$$ \hspace{1cm} (2.5)

so the active portfolio has weights $v_i$ and a constant proportional fee $f_a$ on the portfolio return, while the market portfolio is simply a value-weighted average of individual stock returns. We can also decompose the active portfolio return into $\tilde{R}_a = \alpha_a + \beta_a \tilde{R}_m + \tilde{\varepsilon}_a$ where $\beta_a$ is the market beta of the portfolio and $\tilde{\varepsilon}_a \sim N(0, \sigma_a^2)$. Then the after-fees abnormal return $\alpha_a$ and

66
the idiosyncratic variance $\sigma_a^2$ of the manager’s portfolio are given by:

$$\alpha_a = \sum_{i=1}^{N_S} \nu_i \alpha_i - f_a$$  \hfill (2.6)

$$\sigma_a^2 = \sum_{i=1}^{N_S} \nu_i^2 \sigma_i^2$$  \hfill (2.7)

where $\alpha_i$ and $\sigma_i^2$ denote the abnormal return and the idiosyncratic variance of return for stock $i$.

We assume the end investor knows the expected returns and variances on the active portfolio and the passive market portfolio (but not on individual stocks). These are summary statistics of the stock market which can be learned over time in a repeated-game setting, whereas the alpha of an individual stock is randomly drawn each period and thus cannot be learned over time. The optimal allocations to the active and passive managers are then given by:

$$W_a^* = \frac{E \left[ \tilde{R}_a \right] - \beta_a \eta}{\gamma e \sigma_a^2} = \frac{\alpha_a}{\gamma e \sigma_a^2}$$  \hfill (2.8)

$$W_p^* = \frac{E \left[ \tilde{R}_m \right] - \beta_a W_a^*}{\gamma e \sigma_m^2} = \frac{\eta}{\gamma e \sigma_m^2} - \beta_a W_a^*,$$  \hfill (2.9)

where $\eta$ denotes the market risk premium.

The value function of the investor can be defined as the solution to the maximization problem (2.3). Once we plug in the optimal allocations $W_a = W_a^*$ and $W_p = W_p^*$, that value function can be transformed into a more convenient form:

$$V = \frac{1}{\gamma e} \left[ \left( \frac{\eta}{\sigma_m} \right)^2 + \left( \frac{\alpha_a}{\sigma_a} \right)^2 \right].$$  \hfill (2.10)

Given his optimal portfolio allocations $W_a^*$ and $W_p^*$, the investor’s expected utility therefore depends on the Sharpe ratio of the market and the appraisal ratio of the active portfolio. The appraisal ratio can be interpreted as a measure of the mean-variance efficiency of abnormal
returns, i.e. when we plot alphas against idiosyncratic risk. Hence, the end investors would want the active manager to go for mean-variance efficient portfolios, which is also consistent with the advice of Treynor and Black (1973).

**Active Managers**

So far we have derived the end investors' asset allocations and expected utility for an exogenously given actively managed portfolio. How should the active managers pick this portfolio?

When there are no agency issues involved, probably the most plausible answer is that the manager picks the same portfolio weights the end investor would pick if he could invest in the stock market himself (using the manager's information). These are also the first-best portfolio weights from a contracting problem between the two, as it allows the investor to extract the greatest utility gain from the stock portfolio.

However, when the dollar fee of the manager depends on the end investor's allocation, it is not so clear that the above intuition carries through. Hence, we proceed to derive the active manager's portfolio weights more formally.

The manager receives a proportional fee for the wealth that the end investor allocates to this portfolio. This is what we observe in practice, and in fact it would be very difficult to maintain any other kind of fee structure in equilibrium. Since portfolios are virtually costless to repackage, any nonlinear pricing (including nonlinear fees) would represent an arbitrage opportunity.\(^{10}\)

We allow the active manager to take short positions as well, so the cost of his stock portfolio could be zero or even negative. Yet in reality investors cannot take arbitrarily large long-short positions as they are constrained by a collateral requirement on the short positions.\(^{11}\) We therefore set the cost of short positions equal to zero, so the cost of the portfolio is determined by its long positions only:

\[
\sum_{v_i>0} v_i = 1. \tag{2.11}
\]

\(^{10}\)This concerns the linearity of dollar fees in assets under management. We abstract away from return-based incentive fees since 98% of U.S. mutual funds do not have such fees.

\(^{11}\)Investors are required to deposit 102% of the cash proceeds of the short sale with their broker (D’Avolio (2002)). In the model this collateral requirement acts only as a normalization and as a tool to make larger positions cost more, so its precise form does not matter.
We assume a management fee of $f$ percent of the combined size of the long position and the short position.\footnote{The active manager's portfolio will actually look very much like that of a long-short equity hedge fund. In reality, the fee for these funds is typically around 1.5\% of the cost of the portfolio (i.e. long-only holdings), but there are also significant contingency fees of around 20\% for performance above a benchmark. On the other hand, actively managed mutual funds typically charge proportional fees close to 1.5\% of the portfolio, but these funds are almost exclusively long-only and their returns are benchmarked against the market; hence a fund may effectively end up charging investors also for its investment in the market portfolio. For our purposes, we believe a proportional fee for both the long and the short position is an appropriate compromise. In the extreme case of a fee for the long-only position combined with no contingency fee, the effects observed for a given percentage fee would be approximately cut in half.} The dollar fee is thus given by

$$f \sum_{i=1}^{N_S} |W_a v_i| = f W_a \sum_{i=1}^{N_S} |v_i|,$$

(2.12)

which translates to a fee of

$$f_a = f \sum_{i=1}^{N_S} |v_i|$$

(2.13)

as a fraction of the portfolio investment $W_a$.

We assume that there is a market for active managers. Anyone can become an active manager by paying a fixed dollar cost $C$. This reflects the costs of information acquisition, as well as any fixed costs of being an active trader in individual stocks. It allows the manager to learn the stock-specific parameters $a_i$, $b_i$, and $\sigma_{e_i}^2$ and then actively pick an efficient portfolio with weights $v_i$.

Active managers compete with one another to provide the end investor with a portfolio that maximizes his expected utility (2.10), subject to the constraint that the managers have to earn their costs at the end investor’s optimal allocation $W_a = W_a^*$. Since the active managers take the market risk premium and market volatility as given, maximizing the end investor’s expected utility is equivalent to maximizing the appraisal ratio of the active portfolio. The manager’s problem is then:

$$\max_{\{\{v_i\}, f\}} \frac{\alpha_a}{\sigma_a}$$

s.t. $f W_a^* \sum_{i=1}^{N_S} |v_i| \geq C.$

(2.14)
Substituting in $\alpha_a$, $\sigma_a$, and $W^*_a$ allows us to rewrite the manager’s maximization problem as:

$$\max_{\{v_i\}, f} \frac{\sum_{i=1}^{N_a} v_i \alpha_i - f \sum_{i=1}^{N_a} |v_i|}{\sqrt{\sum_{i=1}^{N_a} v_i^2 \sigma_i^2}}$$

subject to

$$f \frac{\sum_{i=1}^{N_a} v_i \alpha_i - f \sum_{i=1}^{N_a} |v_i|}{\gamma_c \sum_{i=1}^{N_a} v_i^2 \sigma_i^2} \sum_{i=1}^{N_a} |v_i| \geq C.$$

(2.15)

After some algebra, we find that the manager’s optimal portfolio weights are linear in alpha:

$$v_i = \left( \frac{1}{\sum_{\alpha_j > 0} \frac{\alpha_j}{\sigma_j^2}} \right) \frac{\alpha_i}{\sigma_i^2}.$$  

(2.16)

When there is no management fee, we can immediately obtain this result using a CARA-normal setup or mean-variance analysis. However, obtaining the same result as a solution to (2.14) confirms that our proportional fee does not change the optimal portfolio weights chosen by the manager.\(^\text{13}\)

In our model of competitive active money managers, we have a fixed dollar cost but no diseconomies of scale. Therefore, in equilibrium with free entry, there will be only one active manager whose total fees are exactly enough to cover his fixed dollar cost $C$. If the manager’s fees exceed his cost, someone else will step in, undercut the fees of the incumbent, and win the business of all end investors. In reality we of course observe a large number of competing yet coexisting actively managed funds and fund families, presumably due to some organizational diseconomies of scale. While it would certainly be realistic to include these considerations in our model, it might also shift the focus away from the main object of interest in this chapter, namely the effect of the intermediaries and their proportional fee on the cross-sectional pricing of assets. Hence, we view our simple setup primarily as evidence that the intuitively appealing mean-variance portfolio weights can also be formally justified.

The dollar demand of the active manager for stock $i$ can then be expressed as

$$W_i = W_a v_i = \frac{W_a \alpha_i}{\sum_{\alpha_j > 0} \frac{\alpha_j}{\sigma_j^2}} = \frac{\alpha_i}{\gamma \sigma_i^2},$$

(2.17)

\(^{13}\)Section 5.1 further discusses the optimality of this contract.
where we defined the “effective risk aversion” of the active manager as

$$\gamma = \frac{1}{W_0} \sum_{\alpha_j > 0} \frac{\alpha_j}{\sigma_j^2}. \quad (2.18)$$

This is the implied coefficient of absolute risk aversion of the active manager if he was a CARA investor investing his own wealth. Since in reality the manager simply invests all his assets under management in stocks, his effective risk aversion is directly determined by the end investor’s dollar allocation to him. This notation not only simplifies our equations but it also offers a convenient interpretation in the equilibrium analysis.

**Equilibrium**

There are three groups of investors holding stock $i$: First, the passive manager holds the same fraction $u_p = \frac{W_p}{P_m}$ of the supply of each stock, where $P_m$ is the price of the market portfolio. His demand will therefore depend not on the price of stock $i$ but on the price of the aggregate market portfolio. Second, noise traders hold a random supply $u_{in} \sim N(0, \sigma_u^2)$ which is independent of price. These are the investors who create profitable trading opportunities for sophisticated stock pickers. Third, the active manager holds a supply $u_i$. Thus it is the active manager whose actions will determine the cross-sectional pricing of stocks. Together, the demand of the three investors adds up to the supply of the stock:

$$u_p + u_{in} + u_i = 1. \quad (2.19)$$

The equilibrium price of the stock will be

$$P_i = \frac{\alpha_i}{b_i \eta} - \frac{\gamma \sigma_{ei}^2 u_i}{P_i}. \quad (2.20)$$

This yields an alpha of

$$\alpha_i = \frac{\gamma \sigma_{ei}^2}{P_i} u_i. \quad (2.21)$$

By construction, the market portfolio will always have alpha of zero. This implies that $u_i \sim N(0, \sigma_u^2)$. In other words, the active manager will hold an equal number of shares in his long
and short positions, so his exposure to market risk will automatically be zero.

We then have five remaining equilibrium variables: the allocations $W_a$ and $W_p$ to the active and passive managers, the market risk premium $\eta$, as well as the fee $f$ and the effective risk aversion $\gamma$ of the active manager. We also have five equations: two for the allocations, one for the portfolio value of the active manager, one for the market clearing of stock $i$, and one for the dollar fee. After some algebra, we obtain the following:

**Proposition 1** The equilibrium is given by:

\[
\eta = \frac{\gamma_e \sigma^2_M}{N_S a - \gamma_e \sigma^2_M} \quad (2.22)
\]

\[
W_p = \frac{N_S a - \gamma_e \sigma^2_M}{N_S a - \gamma_e \sigma^2 u} \quad (2.23)
\]

\[
W_a = \frac{N_u \sigma_u}{2} \left[ \frac{\sqrt{\frac{2}{\pi}} (a - b \eta) - \gamma_e \sigma^2 u}{\sigma^2 u} \right] \quad (2.24)
\]

\[
f = \frac{\sqrt{\frac{2}{\pi}} (a - b \eta)}{N_S \sigma_u} \quad (2.25)
\]

\[
\gamma = \gamma_e + \sqrt{\frac{2}{\pi}} \left( \frac{a - b \eta}{\sigma^2 u} \right) f. \quad (2.26)
\]

Here $\sigma^2_M$ denotes the dollar variance of the market portfolio. We leave some expressions in terms of the market risk premium $\eta$ to keep them simple, and we leave the last expression in terms of the endogenous variable $f$ as we prefer to calibrate the model to a percentage fee rather than a dollar cost.

### 2.3.3 Analysis of Equilibrium

To calibrate the model, we set the length of the period to one year, the number of stocks $N_S = 1,000$, the risk aversion of the end investors $\gamma_e = 1.5625 \times 10^{-8}$ (to produce a market risk premium of $\eta = 0.05$), $a = 105$ (to normalize the average price to 100), $b = 100$ (to set the beta of the market portfolio $\beta_m = 1$), $\sigma^2_M = 4 \times 10^8$ (to get a standard deviation of 20% for the market return), $\sigma^2 u = 900$ (to get a standard deviation of 30% for idiosyncratic stock return), and the dispersion in noise trader holdings $\sigma_u = 0.1$ (so that the 95% confidence interval for noise trader holdings is 40% of the supply of the stock).

The interesting part of the equilibrium is the value of the effective risk aversion of the active
manager:

$$\gamma = \gamma_e + \sqrt{\frac{2}{\pi}} \left( \frac{a - bn}{\sigma^2 \sigma_u} \right) f. \quad (2.27)$$

If the fee $f$ charged by the active manager is zero, then the active manager's risk aversion will match that of the representative end investor. Consequently, a $-10\%$ supply shock to a typical stock will increase the price of the stock by only 0.11 basis points, just like in the simple CAPM calibration we did earlier. However, the fee $f$ has a very significant first-order effect on $\gamma$ - even a tiny fee of 0.1% would increase $\gamma$ by a factor of 70. Table 2.1 illustrates the effect of the fee on the equilibrium distribution of alphas, on the effective risk aversion, and on the price impact of a $-10\%$ supply shock (which would correspond to a stock being added to the S&P 500).

<table>
<thead>
<tr>
<th>fee</th>
<th>95% confidence interval for $\alpha_i$</th>
<th>effective risk aversion $\gamma$</th>
<th>price impact of a $-10%$ supply shock</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$[-0.0022%, 0.0023%]$</td>
<td>$1.25 \times 10^{-5}$</td>
<td>0.0011%</td>
</tr>
<tr>
<td>0.1%</td>
<td>$[-0.16%, 0.16%]$</td>
<td>$8.99 \times 10^{-4}$</td>
<td>0.081%</td>
</tr>
<tr>
<td>0.5%</td>
<td>$[-0.79%, 0.81%]$</td>
<td>$4.45 \times 10^{-3}$</td>
<td>0.40%</td>
</tr>
<tr>
<td>1.0%</td>
<td>$[-1.6%, 1.6%]$</td>
<td>$8.88 \times 10^{-3}$</td>
<td>0.80%</td>
</tr>
<tr>
<td>1.5%</td>
<td>$[-2.3%, 2.5%]$</td>
<td>$1.33 \times 10^{-2}$</td>
<td>1.20%</td>
</tr>
<tr>
<td>2.0%</td>
<td>$[-3.1%, 3.3%]$</td>
<td>$1.77 \times 10^{-2}$</td>
<td>1.60%</td>
</tr>
</tbody>
</table>

Table 2.1: The effect of the management fee; one-year horizon.

For a realistic fee of 1.5% of assets under management, we get a price impact of 1.20%. This is some orders of magnitude (about 1,000 times) greater than in the classical CAPM case with a zero fee. For even very small values of the fee (0.1%), the risk aversion of the end investors actually becomes irrelevant to the effective risk aversion of the active manager.

This is in stark contrast to traditional representative agent models where end investors' risk aversion will show up both in the pricing of market risk and in the pricing of idiosyncratic risk. In our setup, no such link exists. The market portfolio is still priced according to the risk aversion of the end investors, but the cross-sectional pricing of stocks is determined separately by the fee charged by the professional stock pickers.

What exactly is driving this result? In equilibrium, the end investors will allocate wealth to
the active manager only if he delivers a satisfactory return net of fees. The alpha of his mean-variance efficient portfolio will therefore have to be slightly greater than the fee. The portfolio alpha then determines the equilibrium distribution of stock alphas, as the alphas of individual stocks will have to be sufficiently dispersed to produce a portfolio alpha slightly above the fee.

The slope of the active manager’s demand curve and the distribution of alphas with respect to his information set are closely linked. If the slope is close to zero, his demand will be very elastic and his perceived distribution of alphas will be highly concentrated around zero. If the slope is large, his demand will be less elastic and he will perceive a more dispersed distribution of alphas. Hence, the model effectively determines an equilibrium level of inefficiency in the market, measured with respect to the active manager’s information set.

While our explanation produces an enormous increase in price impact relative to its neoclassical benchmark, it still falls short of the 13% price impact observed for the S&P 500. However, there are two main reasons why our numbers should only be considered a lower bound resulting from the effect we described. First, we conducted the analysis in a one-period setting with fixed payoffs one year later, so we implicitly assumed that all stock prices will return to their fundamental values in a year. In a more realistic infinite-horizon setting there is no such guaranteed convergence (in fact many people believe the price premium following S&P 500 index addition will not disappear as long as the stock remains in the index), and therefore we would expect a greater price impact today. To illustrate this effect, we next calibrate our model to a five-year horizon. Second, we assumed costless short-selling by the active manager. In reality, most active investors face short-sales costs or constraints which can have a significant effect on the supply available to a new buyer. A more elaborate version of our model in the appendix will address this second issue.

Table 2.2 shows the effect of a longer (five-year) horizon on our results. The speed of convergence turns out to matter a great deal for stock prices today. This is because investors care about the alpha per unit time as the fee is also charged per unit time, and when prices converge over a longer period of time, the cumulative alpha over the entire period is capitalized into the stock price today.

The five-year horizon can be interpreted as implying a half-life of 2.5 years for all “mispricings.” This seems fairly plausible in itself, and it would be consistent for example with the
Table 2.2: The effect of the management fee; five-year horizon.

<table>
<thead>
<tr>
<th>annual fee</th>
<th>95% CI: cumulative $\alpha_i$ over 5 years</th>
<th>effective risk aversion $\gamma$</th>
<th>price impact of a $-10%$ supply shock</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$[-0.012%, 0.012%]$</td>
<td>$1.25 \times 10^{-5}$</td>
<td>$0.0062%$</td>
</tr>
<tr>
<td>0.1%</td>
<td>$[-0.80%, 0.82%]$</td>
<td>$8.15 \times 10^{-4}$</td>
<td>$0.41%$</td>
</tr>
<tr>
<td>0.5%</td>
<td>$[-3.8%, 4.2%]$</td>
<td>$4.02 \times 10^{-3}$</td>
<td>$2.0%$</td>
</tr>
<tr>
<td>1.0%</td>
<td>$[-7.4%, 8.7%]$</td>
<td>$8.04 \times 10^{-3}$</td>
<td>$4.0%$</td>
</tr>
<tr>
<td>1.5%</td>
<td>$[-11%, 14%]$</td>
<td>$1.20 \times 10^{-2}$</td>
<td>$6.0%$</td>
</tr>
<tr>
<td>2.0%</td>
<td>$[-14%, 19%]$</td>
<td>$1.61 \times 10^{-2}$</td>
<td>$8.0%$</td>
</tr>
</tbody>
</table>

Empirical evidence of slow mean reversion over a 3-5-year period (DeBondt and Thaler (1985)) as well as the slow return reversals documented in studies on short-term momentum (Jegadeesh and Titman (1993)). If we wish to consider the index premium as a permanent price effect which is not reversed until the stock is deleted from the index, then we could interpret the five-year horizon for an index addition as the expected lifetime of the stock in the index. In contrast, it would be very hard to find similar support for one-year full reversal of mispricings – we have used the one-year horizon in this chapter only as a first cut because it makes the model’s parameter values and predictions a little more transparent.

Naturally we are not purporting to account for everything that matters for the slope of the demand curve, so we should not be surprised to get a smaller value than empirically observed. For a 1.5% fee and a $-10\%$ supply shock, our model predicts a price impact of about half of the actual price premium for S&P 500 additions, which shows that we get the correct order of magnitude. But for the aforementioned reasons, the predicted price impact can be even greater if we cast our economic story in a more realistic setting.

### 2.3.4 Empirical Implications

Most of the model’s testable implications stem from two equations:

\[ P_i = a_i - b_i \eta - \gamma \sigma^2_{e_i} u_i \]  
\[ \gamma = \gamma_e + \sqrt{\frac{2}{\pi}} \left( \frac{a - b\eta}{\sigma^2_{e} \sigma_u} \right) f. \]  

\[ (2.28) \]  
\[ (2.29) \]
The price of a stock is given by its CAPM price \((a_i - b_i \eta)\) minus a deviation \((\gamma \sigma^2_i u_i)\) due to idiosyncratic risk. As the equilibrium holdings \((u_i)\) of the active manager change, the price impact is given by the dollar variance \((\sigma^2_i)\) of the stock's payoff times the effective risk aversion \((\gamma)\) of the active manager. The price elasticity of demand for stock \(i\) is then

\[
\frac{dQ_i}{dP_i} = \frac{du_i}{dP_i} = P_i \frac{d\sigma_i^2}{dP_i} = \frac{-P_i}{\gamma \sigma^2_i}.
\]

(2.30)

**Implication 1** The demand curve is steeper for stocks with greater idiosyncratic risk.

The effective risk aversion of the active manager is supposed to be the same across all stocks. However, if the stock market is segmented so that each active manager (stock picker) generally focuses on a subset of the available stocks, we may also see some variation in the manager’s effective risk aversion as his fee changes from one segment to another.

**Implication 2** The demand curve is steeper for stocks in segments of the market with a greater fee for active management.

**Implication 3** The demand curve is steeper for stocks in segments of the market with a greater cost of information acquisition.

The latter implication holds when the fee for active management is related to the information acquisition cost of the manager.

**Implication 4** The demand curve is steeper for stocks in segments of the market with less dispersion in noise trader holdings.

It may be somewhat surprising that a larger dispersion of noise trader holdings actually makes demand curves more horizontal and in that sense makes the market more efficient. The reason is that the equilibrium dispersion of alphas across stocks would have to be the same as the active managers still earn their fees, but now the same dispersion of alphas would be observed over a wider range of the managers’ stock holdings, so the change in alpha (and price) for a supply shock of a given size would be smaller.

Our model also implies that noise traders can move prices, and in fact they can increase the volatility of a stock beyond the volatility of its fundamentals.
Implication 5  Stocks with a greater volatility of noise trader holdings will exhibit greater price volatility, unless the shocks to noise trader holdings are inversely correlated with fundamental news.

2.4 Empirical Tests

The most unique implications of our model are perhaps Implication 2 and Implication 3, since they link the fee of the intermediary to the slope of the demand curve. However, it is very hard to construct a clean test of this link.

One possible test using U.S. stock market data is whether small-cap stocks have steeper demand curves because the fees of active small-cap money managers are generally higher than the fees of large-cap managers, presumably reflecting the higher information acquisition costs for smaller firms. We can compare the demand curves implied by index changes for the Russell 2000, which is a small-cap index, with the large-cap S&P 500. Of course this is far from a clean test since it is not easy to control for all other relevant differences between the two indices. Nevertheless, our results in Chapter 1 reveal that the demand curves for Russell 2000 stocks indeed seem to be steeper than those for S&P 500 stocks, by about a factor of 3 or 4. Hence, this result is consistent with the main prediction of our model.

The link of Implication 1 between idiosyncratic risk and the demand curve slope provides an opportunity for a relatively straightforward test. In Chapter 1, we performed this test both for the Russell 2000 and for the S&P 500. The Russell 2000 provides a much larger sample of event stocks than the S&P 500 with hundreds as opposed to tens of stock each year. The Russell 2000 data are also relatively untouched by academic researchers, as almost all U.S. index studies have focused on the S&P 500. We are able to confirm the prediction for both data sets: idiosyncratic risk has both a statistically and economically significant impact on the slope of the demand curve for a stock.
2.5 Interpretations and Further Discussion

2.5.1 Optimality of Active Managers’ Policy

Even though we derived the policy of an active manager as a solution that maximizes the end investors’ utility, subject to the manager’s break-even constraint, we obtained a result that is very different from the CAPM benchmark. If the end investors had the managers’ information themselves, they would invest about 1,000 times as aggressively as they will through intermediaries with a 1.5% fee. Clearly the results are driven by the fee, but what exactly is it that makes the effect so large?

Regardless of the structure of competition between active managers, the first-best contract between an active manager and an end investor (when both act as price takers) would include a lump-sum dollar payment to the manager in exchange for allowing the investor to invest freely in the before-fees mean-variance efficient portfolio. In other words, the management fee would not interfere with the investor’s portfolio selection. Given the considerable dispersion in equilibrium alphas that we observe for a 1.5% fee, an end investor and a manager should agree that the end investor gets to invest as much as he likes in the mean-variance efficient portfolio with zero marginal cost, which would make him invest extremely aggressively, while the manager would receive a relatively large lump-sum dollar fee.

However, this first-best contract cannot be implemented for a very fundamental reason. The manager cannot plausibly verify each end investor’s risk aversion, nor can he prevent end investors from pooling their portfolios to reduce the fee per investor. The manager can infer these quantities ex post from the end investor’s portfolio choice, but if he then adjusts his lump-sum dollar fee, we are back in the world of linear management fees. In fact, since repackaging portfolios is virtually costless, the absence of arbitrage makes the dollar management fee linear in the dollar size of the portfolio. This is what we also observe in practice.\(^{14}\)

When an end investor faces a linear dollar fee, he will dramatically scale back his investment to minimize the fee paid to the manager. In equilibrium, the distribution of alphas will be sufficiently wide to allow the active manager to earn his proportional fee, but any abnormal

\(^{14}\)While mutual fund expense ratios are constant proportional fees, private investment advisors may in fact charge a proportional fee where initially the percentage fee slightly declines with portfolio size and then stays constant above a certain size. This could reflect the fixed costs of personally managing an investor’s account.
performance above the fee will again be aggressively exploited by the end investor. Hence, it
is the inefficient but unavoidable linearity of the management fee that makes the end investors
invest so conservatively through the intermediaries.

2.5.2 Relationship to Grossman and Stiglitz (1980)

Our basic economic story with an “equilibrium degree of disequilibrium” is very much in the
spirit of the insightful paper by Grossman and Stiglitz (1980). Could we perhaps use their model
or a multi-asset generalization of their model to explain downward-sloping demand curves?

Grossman and Stiglitz present a single-asset model with informed investors, uninformed
investors, and noise traders. The informed traders observe a signal of the fundamental value of
the asset. The uninformed investors use the price of the asset to infer the signal of the informed,
but the inference is noisy due to the unobserved holdings of noise traders. An uninformed
investor can also become informed by paying a certain cost. The fraction of investors who
choose to become informed is determined endogenously, so that in equilibrium the investors are
indifferent between the two choices. The cost of becoming informed determines the equilibrium
level of “inefficiency” in the market.

Part of the reason demand curves slope down in that model is that the uninformed investors
cannot distinguish whether a supply shock came from the informed traders (because they re-
ceived good news about the stock) or the noise traders (conveying no information about the
stock). However, we are concerned about demand curves for stocks in the absence of new
information. For example, when a stock is added to the S&P 500, every active trader in the
stock who is not consciously ignoring news will know who the new buyers are and why the stock
price went up. Thus any price effect from index addition would have to come from the risk
aversion of the investors and not the rational expectations story of the model.

Extending Grossman and Stiglitz to multiple assets (the entire stock market) would create
a very large unconditional dispersion of the payoffs due to the large differences in the operating
sizes of firms. In the entire cross-section of stocks the uninformed investors can no longer use
the market value of a firm to infer much about its expected return because those market values
can easily vary across stocks by a factor of 1,000.\textsuperscript{15} Even the variation in price-to-book or

\textsuperscript{15}See Berk (1995) for a discussion of similar issues. We also build a model where we explicitly deal with wide
price-to-earnings ratios in the cross-section of stocks is so high that forming portfolios based on those ratios will produce almost passive strategies in the time series of a single stock. Hence, in a multi-asset extension of the model, the uninformed investors would become essentially passive investors who hold the market portfolio (or perhaps a portfolio with more weight on low price-to-book stocks). This would render their model setup closer to ours: it would now have informed active investors, uninformed passive investors, and noise traders.

To generate the same slope for the demand curve as in our model with a fee of 1.5%, the representative informed investor would have to have a risk aversion equal to the effective risk aversion of our active manager (see Table 2.2 on page 75). This implies that one investor out of 1,000 would choose to become informed. If an investor accounts for 0.1% of the market when he is uninformed, the same investor would hold a long-short position with a combined value of close to 10% of the market once he has become informed. It seems like a stretch to say that this enormous increase in his risky portfolio came from the investor’s personal wealth (or personal borrowing which would require collateral) – instead we could interpret this more plausibly as the investor becoming an informed intermediary who primarily invests other people’s money.

But once the investor starts investing other people’s money, we can no longer use his personal risk aversion to explain his investment behavior! His effective risk aversion would now be determined by how much wealth the other investors are willing to allocate to him. Yet the model effectively assumes even the informed investors still keep investing their own wealth but they just borrow massively from outside of the model to finance virtually all of their new and very large portfolios. Thus the model is missing the crucial part of the mechanism which is the tradeoff of end investors when allocating wealth to active managers and the resulting equilibrium value for the effective risk aversion of the active managers.

Of course this is not a deficiency of the original model of Grossman and Stiglitz, since it primarily illustrates rational expectations and information acquisition in a single-asset context as opposed to describing the asset allocation in the entire stock market. We simply want to point out that in a general equilibrium setting, it is essential to explicitly model the delegation of portfolio management if we want to build a plausible model with costly information acquisition.

dispersion in operating sizes of firms.
2.5.3 Institutions Make Prices More Efficient

Our institutional setup seems to arise naturally in response to the underlying fixed cost for active portfolio management. But what if we still want to eliminate the institutions from the model, for a moment ignoring our previous concerns about individuals only investing their personal wealth? Then we will be left with the basic CAPM setting, except that some end investors pay the fixed cost and share it among one another, while others do not pay anything and only invest in the market portfolio. What will be the impact of the underlying fixed cost in these two settings?

The cross-sectional pricing of stocks depends on the (effective) risk aversion of those investors who have paid the cost. Without institutions, we find that in equilibrium only a small fraction $\mu$ of the end investors pay the cost:

$$\mu = \frac{\gamma e N_S \sigma^2 \sigma_u^2}{2C}. \tag{2.31}$$

This means that the risk aversion of those active investors collectively will be

$$\gamma = \max \left\{ \gamma e, \frac{2C}{N_S \sigma^2 \sigma_u^2} \right\}. \tag{2.32}$$

In the institutional setting, the effective risk aversion of the active manager is

$$\gamma = \gamma e + \frac{C}{N_S \sigma^2 \sigma_u^2}. \tag{2.33}$$

Hence, the institutional setting cuts the risk aversion of the active investors by about one half compared with the setting without institutions. The institutions are therefore not a source of friction in the model. Instead, the true source of friction is the underlying fixed cost, and the presence of institutions actually mitigates the effect arising from that cost. This is because the institutions allow all investors to share some of the risk of the actively managed portfolio, while the proportional fee makes it individually optimal for investors not to invest so aggressively that they would collectively drive the alphas to zero.

Since the active investors are less risk-averse in the institutional setting, demand curves will be flatter and prices will be closer to the neoclassical CAPM benchmark. Thus the institutions
emerge to reduce the underlying friction of the model, and the institutions in fact bring prices closer to their fundamental values. In other words, they make the market more informationally efficient.

2.5.4 Transaction Costs

Could we perhaps interpret the management fee in our model as a transaction cost that the representative investor has to pay when trading individual stocks? Would this produce results similar to our setup with financial intermediaries?

The first challenge for transaction costs is their magnitude. Stocks added to the S&P 500 typically have a market capitalization of several billion dollars. Transaction costs for turning around a position in such mid-cap and large-cap stocks are likely to be a fraction of a percent. Yet the S&P 500 premium is about 13%, which certainly seems sufficient to produce abnormal returns even net of transaction costs. Moreover, the largest additions such as Goldman Sachs, UPS, and Microsoft have the lowest transaction costs, yet they tend to experience the largest price impact. Our empirical results for the S&P 500 suggest that larger firms (with lower transaction costs) tend to have steeper demand curves, although the difference is not statistically significant, while the transaction cost story would suggest the exact opposite.

More fundamentally, when end investors trade stocks themselves, they will very aggressively exploit any alphas net of transaction costs, again due to the low risk aversion implied by the market risk premium, so that in equilibrium such abnormal returns cannot exist. Yet downward-sloping demand curves imply that prices (and alphas) change relatively smoothly as we vary the size of the supply shock. There are two ways in which transaction costs could generate demand curves similar to the ones in our intermediary setting: First, if the transaction cost for a stock moves synchronously with alpha, the net-of-costs alphas can stay constant. Second, if the transaction cost is constant but it somehow excludes everyone except one in a thousand investors from trading, and that fraction is more or less independent of the stock price, then the effective risk aversion of the end investors increases to what we had for a 1.5% fee and we again obtain downward-sloping demand curves.\footnote{Introducing heterogeneity into the beliefs of investors would not help much in justifying this setup. In equilibrium the end investors would be able to disagree about the value of a stock only within the narrow bands} Needless to say, this means

82
that transaction costs cannot account for the results of our model.

Of course this does not prove that transaction costs cannot be a plausible explanation in a more complicated setup. The challenge here would be to find that missing ingredient in a model that can allow realistic values of transaction costs to produce economically significant slopes for the demand curves.

### 2.5.5 CRRA Utility

In a multi-asset setup, the normal distribution for stock returns combined with CARA utility offers by far the greatest analytical convenience. However, to verify that our numerical results are not specific to constant absolute risk aversion, we solve for an approximation to constant relative risk aversion in our basic CAPM benchmark. We calibrate the representative investor’s risk aversion by assuming a lognormal distribution for the market return. Yet the idiosyncratic risk of an individual stock has such a small impact on the wealth of the investor that we can still apply CARA analysis locally for a given level of wealth. The possibly meaningful difference with global CARA utility arises from the fact that now the investor has to evaluate idiosyncratic risks for a random local coefficient of absolute risk aversion, where the coefficient depends on the investor’s wealth which in turn depends on the random return on the market portfolio. When the investor has invested all his wealth in the stock market and the market volatility is 20%, the 95% confidence interval for the investor’s local coefficient of absolute risk aversion will be \([1^{4.7}, 6.6]\). We compute the investor’s approximate expected utility and take the first-order conditions to find the demand for each individual stock. The difference with the basic CARA case turns out to be negligible for a one-year horizon. This is perhaps not surprising, because the wealth effects induced by CRRA utility can only show up when there is very large variation in the investor’s future wealth.

### 2.5.6 Interpretation of Long-Short Positions

How should we think about our model in a realistic world where only a small fraction of investors ever take short positions?

of the transaction cost; otherwise they would take extreme positive and negative positions in individual stocks.
In our earlier calibration, the 95% confidence interval for the holdings of both the noise traders and the active managers was \([-20\%, 20\%]\) of the supply of the stock, while the passive manager held exactly 100% of the market portfolio. If we simply shift 20% of the market portfolio to the active managers and 20% to the noise traders, the individual stock positions begin to look more reasonable, as the noise traders and the active managers would short only about 2% of the stocks. The active managers would hold large positions in the market portfolio, but they would also be benchmarked against it and their alpha would still be derived from the long-short portfolio on top of the market portfolio. Moreover, if the managers still charge the same fee of assets under management, in this case the managers’ effective fee for the same long-short portfolio would be increased by a factor of about 2.5 (relative to the earlier calibration). This would scale up the slopes of the demand curves by the same factor, so the effects of the intermediaries would become more prominent — in fact it would turn the predicted 6% price impact to a 15% price impact, which is almost identical to the current S&P index premium.

2.6 Conclusions

In a standard neoclassical multi-asset setting such as the CAPM, both the market risk premium and the slope of the demand curve for an individual stock are jointly determined by the risk aversion of the representative investor. If we back out the representative investor’s risk aversion from any empirically plausible market risk premium, we find a relatively low implied risk aversion; if we back it out from the empirically observed slope of the demand curve for an individual stock, we find a relatively high implied risk aversion. The two estimates differ by several orders of magnitude, presenting us with a well-known puzzle in finance.

In this chapter we propose a possible explanation for the puzzle. In traditional models it is generally assumed that financial intermediaries have no meaningful effect on prices so that we can ignore them and let the owners of wealth invest directly in the stock market. However, this may not always be an appropriate assumption. When most of the active and informed money is controlled by professional active money managers, the slope of the demand curve for a stock is determined by the implied risk aversion of these active managers which, in turn, depends on the wealth allocated to them. Since the active managers charge a fee for their services, their
implied risk aversion in equilibrium can be determined almost entirely by their fee and not by the risk aversion of the end investors. Yet the end investors still set the aggregate market risk premium according to their actual risk aversion.

This result arises from a straightforward intuition: in equilibrium, the active managers have to approximately earn their fees. Thus there persists an equilibrium level of market “inefficiency” exploited by the active managers to recover what are presumably their fixed costs for acquiring information and actively trading on it. The costlier the information, the steeper the demand curves should be.

The magnitude of this effect can be surprisingly large. In our calibration, increasing the fee from zero (the CAPM benchmark) to 1.5% can increase the slope of the demand curve by a factor of 1,000. With a five-year horizon, this fee may increase the price impact of the S&P 500 index membership shock from less than one basis point (CAPM) to a very significant 6.0%. When we allow active managers to hold market risk and be benchmarked against it, as in the real mutual fund industry, the implied price impact increases to 15%.

The steeper demand curves we found in Chapter 1 for Russell 2000 stocks as opposed to S&P 500 stocks provide some supportive evidence for our predicted link between the fee for active management and the slopes of the demand curves for stocks. Our cross-sectional tests in Chapter 1 also confirmed the link between the slope and idiosyncratic risk predicted by our model.

We believe this chapter makes two main contributions. It suggests a generally applicable explanation to the puzzle about downward-sloping demand curves which produces not only the correct sign for the effect but also the correct order of magnitude. More broadly, it provides a concrete example to reaffirm the answer to the title question of Allen's presidential address (2001): Yes, financial institutions do matter, and we do not even need agency issues to generate this result.
2.7 References


Goetzmann, W.N. and M. Massa, 1999, “Index Funds and Stock Market Growth,” working paper, Yale School of Management and INSEAD.


88


89


2.8 Appendix: A More Elaborate Model

2.8.1 Motivation

Our earlier model provides a simple illustration of our economic story. But its frictionless setup also produces the result that prices in the market are set by long-short investors who take large positions each way and whose net portfolio value can even be negative. Yet in reality we observe relatively small short interest for most stocks.

Dechow et al. (2001) report that about 80% of the firm-years in their sample\(^\text{17}\) have a short interest less than 0.5% of shares outstanding, and less than 2% of the firm-years have a short interest greater than 5%. Nowadays short-selling is a little more common, and e.g. for August 15, 2002, the NYSE reported a record short interest of 2.3% of all shares outstanding.\(^\text{18}\) Since this figure includes the shares that were shorted for various hedging motives, the average short interest due to fundamental investors (i.e. stock pickers) is even smaller. This general unwillingness to short stocks could arise at least in part as a consequence of the short-sales costs documented by e.g. Jones and Lamont (2002) and D'Avalio (2002).

When a stock is added to the S&P 500 and mechanical indexers buy about 10% of the shares outstanding, most of the supply seems to come from investors who owned the stock before the event. E.g. for the event of July 19, 2002, when seven large U.S. firms replaced seven non-U.S. firms in the index, the average short interest one month before the event, between the announcement and effective days, and one month after the event were 3.0%, 3.2%, and 5.0%, respectively, for the additions, and 2.6%, 2.8%, and 2.2% for the deletions, while the overall NYSE short interest was 2.2%, 2.1%, and 2.3%.\(^\text{19}\) While this event suggests that about 2% of the required 10% supply came from short sellers, historically the number is likely to be even smaller.

Hence, most of the fundamental stock valuation and stock-picking clearly has been and still seems to be done by long-only investors rather than unconstrained long-short investors. We can accommodate this by changing the interpretation of our simple model as we do in Section 2.5.6, or by building it explicitly into the model as we do here.

We have three main reasons to build a more elaborate (and realistic) version of the model: First, the new version of the model serves as a robustness check on the results of the simple model. Second, numerical calibration is generally easier and more easily interpretable for a more realistic model setup. Third, it turns out this setup can give us some results even if the end investors are not fully rational.

2.8.2 The Model

The basic setup is the same as before. There is a risk-free asset yielding an interest rate of zero, and \(N_S\) stocks with terminal payoffs \(\tilde{z}_i = a_i + b_i \tilde{y}_i + \tilde{c}_i\). The end investors maximize CARA utility by optimally allocating their wealth to active managers, passive managers, and the risk-free asset. Again we abstract entirely from agency issues and let the active managers simply follow the orders they are given.

\(^{17}\) NYSE and AMEX stocks, 1976-1993.


\(^{19}\) Data from the exchanges, published monthly by the Wall Street Journal.
There are essentially five differences with the simple model presented earlier. First, the active managers can only take long positions in stocks. Second, because of this short-sales constraint, the active managers will be benchmarked against the market portfolio. Third, there are multiple active managers and they have heterogeneous beliefs about stocks so that all stocks will be held in equilibrium. Fourth, we allow for wide dispersion in the operating sizes of firms \((a_i)\). Fifth, each active manager will have beliefs about a subset of the stocks but not all of them.

**Assets**

There is enormous dispersion in the market capitalization of firms. If we take only the largest 3,000 stocks (which constitute the Russell 3000 index and still represent less than a half of all stocks listed on the NYSE, AMEX, and Nasdaq) at the end of 2001, we get a distribution of values from about $130 million to $400 billion.

We let the constant \(a_i\) of the payoff of stock \(i\) to be distributed as \(\log(a_i) \sim U(\log(a_{\text{min}}), \log(a_{\text{max}}))\). While a lognormal distribution would fit the data better, we pick this form for analytical tractability. What matters is the degree of dispersion, not its exact shape.

The dispersion in \(a_i\) almost completely eliminates any size effect from the model. If each \(a_i\) had the same value or if their dispersion was very small, then any uninformed investor would be able to earn above-market returns by simply buying the cheaper stocks and shorting the more expensive ones. But when there is large dispersion in the operating size of firms, this simple correlation between market price and expected return is severely diminished, and the uninformed investors will not be able to do better than the market portfolio. The dispersion in \(a_i\) effectively ensures that the uninformed investors cannot become informed by just using some piece of easily available information.

The dispersion in \(a_i\) also creates dispersion in the dollar supply of idiosyncratic risk. If the same investors know about the same stocks, then the smaller stocks will be more aggressively priced and will have more horizontal demand curves. Since most of these properties are relatively constant across stocks, the dispersion in \(a_i\) implies that the number of market participants in each stock and their aggregate risk tolerance are also roughly proportional to \(a_i\). This is why we cannot allow all investors know about all stocks. It is also somewhat similar to Merton (1987).

We assume \(b_i = P_i\) and \(\sigma_{\mu_i}^2 = P_i^2 \sigma_i^2\), so that each stock will always have a market beta \(\beta_i = 1\) and a fixed return variance of \(\sigma_i^2\). These assumptions have a negligible effect on our numerical results but they do make our equations more convenient and intuitive.

**End Investors**

The representative end investor's problem is again

\[
\max_{\{w_s, w_p\}} E \left[ -\exp \left( -\gamma \tilde{W}_1 \right) \right]
\]

subject to

\[
\tilde{W}_1 = W_0 + W_s \tilde{R}_s + W_p \tilde{R}_m,
\]

(2.34)

(2.35)

92
which produces the same optimal allocations to the active and passive managers:

\[
W_a = \frac{E[\tilde{R}_a] - \beta_a \eta}{\gamma \sigma_a^2} = \frac{\alpha_a}{\gamma \sigma_a^2} \quad (2.36)
\]

\[
W_p = \frac{E[\tilde{R}_m] - \beta_a W_a}{\gamma \sigma_m^2} = \frac{\eta}{\gamma \sigma_m^2} - \beta_a W_a. \quad (2.37)
\]

The end investor's allocation to the active managers therefore depends entirely on the alpha \(\alpha_a\) (net of fees) of those managers. Whatever market exposure comes from the active portfolio, the end investor fully hedges this by reducing his position in the passive portfolio.

**Active Managers**

There are \(K\) active money managers who are all identical ex ante. Therefore the end investor will simply diversify his active portfolio allocation equally across all active managers, giving the manager \(k\) an allocation of \(W_k = \frac{W_A}{K}\).

The manager \(k\) has beliefs about \(M\) stocks which are a subset of the \(N_S\) stocks available. Specifically, manager \(k\)'s belief about the payoff \(a_i\) of stock \(i\) is given by \(a_{ik} \sim U(a_i - \Delta a, a_i + \Delta a)\).

For the same reasons as before, we model each manager as a CARA investor with a coefficient of absolute risk aversion \(\gamma\). Without loss of generality, we construct \(N_S\) uncorrelated hybrid securities with payoffs \(z_i = a_i + \tilde{e}_i\) and prices \(P_i = P_i(1 + \eta)\). This determines the dollar demand of the active manager \(k\) for stock \(i\):

\[
W_{ik} = \max \left \{ \frac{1}{\gamma \sigma_i^2} \left[ \frac{a_{ik}}{P_i} - (1 + \eta) \right], 0 \right \}. \quad (2.38)
\]

Hence, his demand is linear in his perceived alpha \(\alpha_{ik} = \frac{a_{ik}}{P_i} - (1 + \eta)\), or zero if the perceived alpha is negative.

This also reveals why short-sales constraints can only exist in the presence of heterogeneous beliefs. By construction, the average alpha perceived by any investor is zero, so the investor will have a positive demand for about half the stocks and a zero demand for the other half. Thus if all investors have homogeneous information and face short-sales constraints, half the stocks will have zero demand and their prices are not determined in equilibrium.

The manager invests all the wealth \(W_k\) under his management in this portfolio, so \(W_k = \sum_{i=1}^{M} W_{ik}\) and hence his effective risk aversion is given by

\[
\gamma = \frac{1}{W_k} \sum_{i=1}^{M} \frac{\alpha_{ik}}{\sigma_i^2}. \quad (2.39)
\]

Since the end investor is effectively benchmarking the manager against the market portfolio by instructing him to focus on abnormal returns, the manager can ignore the market risk of his portfolio and let the end investors offset this on their own by investing less with the passive managers.

We do not constrain the manager to trade only a subset of \(M\) out of the available \(N_S\) stocks. However, the average alpha of a stock is zero by construction, so for all the stocks that the manager has no information about, his expected alpha is zero and thus his optimal demand for such stocks is zero. Unlike in Merton (1987),

93
here the incomplete diversification of the active managers results from a restriction on their information sets and not on an explicit restriction on their investment universe. Nevertheless, the exact degree of diversification by the active managers (such as whether they are diversified beyond 50 stocks) does not play a role in any of our results.

Each active manager charges a fee $f$ as a fraction of assets under management.

**Equilibrium**

We define the equilibrium as the set of prices and allocations such that the active managers have invested all their wealth under management in portfolios with mean-variance efficient abnormal returns, the passive managers have invested all their wealth under management in the value-weighted market portfolio, the end investors are maximizing their expected utility by optimally allocating their wealth between the active managers, passive managers, and the risk-free asset, and the market clears for all stocks.

In equilibrium, stock $i$ will be held by the passive managers who hold a supply of $u_p = \frac{w_p}{m}$, the noise traders who hold a randomly chosen supply of $u_{n} \sim U(0, \Delta u)$, and the active managers who hold the remaining supply which we denote as $u_i$. Market clearing then requires that

$$u_p + u_n + u_i = 1$$

which implies that $u_i \sim U(u_{min}, u_{min} + \Delta u)$ where $u_{min} = 1 - u_p - \Delta u$.

We assume there is a continuum of managers with a measure of $N$, who know about stock $i$. Their total dollar demand for stock $i$ is then

$$W_i = \begin{cases} \frac{a_i + \Delta a_i}{1 + \eta + 2\sigma_i \sqrt{\frac{a_i}{N} \gamma u_i}} \int_0^{\frac{a_i}{P_i} - (1 + \eta)} \frac{N}{2\Delta a_i} \, da & \text{if } P_i \geq \frac{a_i - \Delta a_i}{(1 + \eta)} \\ \frac{a_i + \Delta a_i}{1 + \eta + 2\sigma_i \sqrt{\frac{a_i}{N} \gamma u_i}} \int_{\frac{a_i}{P_i} - (1 + \eta)}^{a_i + \Delta a_i} \frac{N}{2\Delta a_i} \, da & \text{if } P_i < \frac{a_i - \Delta a_i}{(1 + \eta)} \end{cases}$$

In the latter case the price of the stock is below the valuation of even the most pessimistic investor. This is unlikely unless the dispersion in beliefs is very small, so we focus on the latter case where we have both investors who believe the stock has a negative alpha and investors who believe it has a positive alpha.

The price of stock $i$ will then be

$$P_i = \frac{a_i + \Delta a_i}{1 + \eta + 2\sigma_i \sqrt{\frac{a_i}{N} \gamma u_i}} = \frac{a_i (1 + \Delta_i)}{1 + \eta + 2\sigma_i \sqrt{\frac{\Delta a_i^2 u_i}{\lambda_i}}},$$

where we defined the relative dispersion-of-beliefs parameter $\Delta_i = \frac{\Delta a_i}{a_i}$ and the density of informed investors $\lambda_i = \frac{N_i}{a_i}$. This determines the true alpha (i.e., conditional on $a_i$) of stock $i$ as

$$\alpha_i = \frac{1}{1 + \Delta_i} \left[ 2\sigma_i \sqrt{\frac{\Delta_i \gamma u_i}{\lambda_i}} - \Delta_i (1 + \eta) \right].$$

Analogously to the results of e.g. Miller (1977) and Chen, Hong, and Stein (2002), the price of the stock reflects
the valuation \( a_i (1 + \Delta_i) \) of the most optimistic investor. However, this valuation is discounted by \( \eta + 2\sigma_i \sqrt{\frac{\Delta_i^2}{\lambda_i}} u_i \), which is greater than the market risk premium \( \eta \) and which reflects the active investors’ aversion to idiosyncratic risk, so that the average alpha across all stocks is still equal to zero.

The above equations determine the joint distribution of stock prices and alphas as a function of the minimum fraction \( u_{\text{min}} \) of a stock held by the active managers, the effective risk aversion \( \gamma \) of the active managers, and the market risk premium \( \eta \), in addition to some stock-specific constants. They also have to be consistent with the equilibrium allocations of \( W_a \) and \( W_p \) to the active and passive managers. These five variables have to be solved for simultaneously from the following system of five equations:

\[
\begin{align*}
\alpha_m &= 0 \\
W_a &= \frac{\alpha_a}{\gamma \sigma^2_a} \\
W_p &= \frac{\eta}{\gamma \sigma^2_a} - W_a \\
\gamma &= \frac{K W_a}{W_p} \sum_{i=1}^M \frac{\alpha_{ik}}{\sigma^2_i} \\
u_{\text{min}} &= 1 - \frac{W_p}{P_m} - \Delta u
\end{align*}
\]

(2.44)  
(2.45)  
(2.46)  
(2.47)  
(2.48)

Here \( \alpha_m \) denotes the alpha of the value-weighted market portfolio and \( P_m \) is the price of the market portfolio.

To solve this system of equations, we first need to compute several expressions: the average alpha \( \alpha_m \) of the market portfolio, the average alpha \( \alpha_a \) (net of fees) of the active managers, the idiosyncratic variance \( \sigma^2_a \) of the active managers, and the summation \( \sum_{i=1}^M \frac{\alpha_{ik}}{\sigma^2_i} \) for an active manager. These computations do not lend themselves to easy and intuitive economic interpretation. Hence, we solve for the equilibrium numerically.

Intuitively, the equilibrium is established as follows: Assume we start in an equilibrium with some fee \( f \) which determines the equilibrium allocations \( W_a \) and \( W_p \) and the equilibrium distributions of stock prices and alphas. Then suddenly the fee is increased to \( f' \). Now the active managers can no longer earn their fees, so the end investors will reduce their dollar allocation to the active managers. Once the dollar allocation of the active managers decreases, they become less aggressive, permitting a wider equilibrium distribution of alphas (in equation (2.43), decreasing the active managers’ equilibrium holding \( u_i \) while keeping its variation unchanged will increase the dispersion of alphas). This wider distribution of alphas will increase the average alpha of the active managers. Once the average alpha rises to the same level as the new fee \( f' \), a new equilibrium is reached.

Thus, the intuition of our simple model generalizes to the richer and more realistic model. Here the mechanics of the model are more complicated, but in return we get parameter values and predictions that are easier to interpret (the share of wealth controlled by active managers; no unrealistically high values of short interest).

### 2.8.3 Analysis of Equilibrium

As before, we calibrate the model by setting the number of stocks \( N_S = 1,000 \), the risk aversion of the end investors \( \gamma_e = 1.5625 \times 10^{-5} \) (to produce a market risk premium of \( \eta = 0.05 \)), and the dispersion in noise trader
holdings $\Delta u = 0.4$. We also set $\beta_i = 1$ and the standard deviation $\sigma_i = 0.3$ for the idiosyncratic return for all stocks, and the standard deviation $\sigma_m = 0.2$ for the market return.

We pick $a_{\text{min}} = 1$ and $a_{\text{max}} = 688$ so that the average $a_i$ is still equal to 105 (as before) but now there is large dispersion around this mean value. We set the mass of active managers $K = 10$ and we let each active manager know about $M = 100$ stocks. Then the average measure of managers who know about stock $i$ is $N_i = \frac{KM}{NS} = 1$, and we assume this is proportional to the expected payoff $a_i$ which implies a density $\lambda_i = \frac{1}{100}$ of active managers for all stocks. The scaling of the number of managers is of course irrelevant as we do the calculations for a continuum of stock. Finally, we choose the maximum dispersion of beliefs $\Delta \lambda$ for a stock as 20% of the expected payoff $a_i$.

The meaningful free parameters to be picked in the model are the active managers’ fee $f$, the dispersion of beliefs $\Delta \lambda$, and the dispersion of noise traders’ demand $\Delta u$. The model’s restrictions then determine the joint distributions of $u_i$ (the supply held by active managers), $\alpha_i$, and $\alpha_a$ as well as the allocations $W_a$ and $W_p$ to the active and passive managers, the active managers’ effective risk aversion $\gamma$, and most importantly the slope of the demand curve. The calibration results are in Table 2.3.

<table>
<thead>
<tr>
<th>fee</th>
<th>$\frac{W_a}{W_a + W_p}$</th>
<th>$[a_{\text{min}}, a_{\text{max}}]$</th>
<th>effective risk aversion $\gamma$</th>
<th>price impact of a -10% supply shock</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01%</td>
<td>425%</td>
<td>$[-0.52%, 0.51%]$</td>
<td>$1.72 \times 10^{-2}$</td>
<td>0.25%</td>
</tr>
<tr>
<td>0.1%</td>
<td>135%</td>
<td>$[-1.7%, 1.6%]$</td>
<td>$5.42 \times 10^{-3}$</td>
<td>0.80%</td>
</tr>
<tr>
<td>0.5%</td>
<td>61%</td>
<td>$[-3.9%, 3.5%]$</td>
<td>$1.21 \times 10^{-2}$</td>
<td>1.8%</td>
</tr>
<tr>
<td>1.0%</td>
<td>44%</td>
<td>$[-5.8%, 4.9%]$</td>
<td>$1.72 \times 10^{-2}$</td>
<td>2.7%</td>
</tr>
<tr>
<td>1.5%</td>
<td>36%</td>
<td>$[-7.4%, 5.9%]$</td>
<td>$2.12 \times 10^{-2}$</td>
<td>3.3%</td>
</tr>
<tr>
<td>2.0%</td>
<td>32%</td>
<td>$[-8.9%, 6.8%]$</td>
<td>$2.46 \times 10^{-2}$</td>
<td>3.9%</td>
</tr>
</tbody>
</table>

Table 2.3: The effect of the management fee; one-year horizon.

For a realistic cost of 1.5% of assets under management, the end investors would allocate 36% of their stock market wealth to professional stock pickers and 64% to passive strategies. The price impact following a -10% supply shock would be 3.3%, or about 3 times as large as in our simple model. Compared with the CAPM benchmark, the order-of-magnitude difference is still due to the same story as before, i.e. the fact that the costly delegation of portfolio management severs the link between the market risk premium and cross-sectional stock pricing. However, the short-sales constraints in this model give a further nontrivial boost to the slope of the demand curve, although this clearly does not change its order of magnitude.

When the fee of the active managers tends to zero, the price impact does seem to approach zero and the demand curves become close to horizontal. This also shows up as a very aggressive allocation to the active managers. Convergence in this model is complicated by the fact that a very small fee and consequently a very large allocation to the active managers (financed by shorting the passive managers) leads to the active managers’ portfolio becoming more and more like the market portfolio. Hence, the idiosyncratic variance of the portfolio falls at the same time as the alpha of the portfolio falls, partially offsetting the effect from a lower average alpha.
So while the model does approach the simple CAPM case with almost horizontal demand curves as the fee tends to zero, the model produces more interesting predictions for more realistic values of the fee.

The slope of the demand curve will be steeper if we decrease the dispersion in noise trader holdings $\Delta u$ or increase the dispersion in beliefs or active managers $\Delta \lambda$. Since the differences in pricing in the cross-section are distributed over the interval of noise trader holdings $[0, \Delta u]$, a narrower interval will mean that the demand curve will have to be steeper to produce the same equilibrium dispersion in alphas. The dispersion in beliefs $\Delta \lambda$ enters through the breadth-of-ownership intuition of Chen, Hong, and Stein (2002): As the supply available to the active managers decreases towards zero, only the valuation of the most optimistic manager determines the stock price since the others cannot short the stock. As the supply available to the managers then increases from zero and the price starts to fall, a wide dispersion in beliefs means it takes a greater fall in price to induce the same number of managers to jump in and hold a positive position in the stock. Nevertheless, the model is relatively robust to changes in these two parameters.

As before, increasing the horizon from one to five years will roughly multiply the price impacts by five. Thus the magnitude of the actual index premium is not outside the scope of this model.

Even if the end investors are not fully rational, we can still use this model to describe the slope of the demand curve, given some (not perfectly rational) allocations to the active and passive managers. E.g. if the end investors allocate a little over a third of their wealth to professional stock pickers and invest the rest in the market portfolio or in random portfolios, we would get similar results as in the equilibrium with rational end investors and a fee of 1.5%. Demand curves would still slope down because of the delegation of portfolio management, i.e. because the active managers are constrained to invest no more than 100% of their wealth under management and because the end investors determine the market risk premium separately from the cross-sectional pricing. However, the puzzle about the demand curves then becomes a puzzle about why the end investors do not invest more with active managers who earn positive alphas. The introduction of the fee for active management can provide a rational explanation for this asset allocation puzzle.
2.9 Appendix: Derivation of Formulas

2.9.1 CAPM Benchmark

We write the representative CARA investor’s problem in the mean-variance form as

$$\max_{\{\theta_i\}} E \left[ \tilde{\pi} \right] - \frac{1}{2} \gamma Var \left[ \tilde{\pi} \right]$$

s.t. \( \tilde{\pi} = W_0 + \sum_{i=1}^{N_S} \theta_i (\bar{x}_i - \bar{P}_i) \). \hspace{1cm} (2.49)

Plugging in the budget constraint and the payoff \( \bar{x}_i = a_i + b_i \tilde{y} + \bar{e}_i \) of stock \( i \), we can express the objective function as

$$\max_{\{\theta_i\}} \sum_{i=1}^{N_S} \theta_i (a_i - \bar{P}_i) - \frac{1}{2} \gamma \left( \sum_{i=1}^{N_S} \theta_i b_i \right)^2 \sigma_m^2 - \frac{1}{2} \gamma \sum_{i=1}^{N_S} \theta_i^2 \sigma_{\bar{e}_i}^2.$$ \hspace{1cm} (2.50)

The first-order condition with respect to \( \theta_i \) is then given by

$$a_i - \bar{P}_i - \gamma \left( \sum_{j=1}^{N_S} \theta_j b_j \right) b_i \sigma_m^2 - \gamma \sigma_{\bar{e}_i}^2 \theta_i = 0.$$ \hspace{1cm} (2.51)

In equilibrium, the investor holds the available supply \( \theta_i = u_i \) of stock \( i \), which determines the stock price:

$$\bar{P}_i = a_i - \gamma \left[ \sigma_m^2 \left( \sum_{j \neq i} u_j b_j \right) b_i + \left( \sigma_m^2 b_i^2 + \sigma_{\bar{e}_i}^2 \right) u_i \right],$$ \hspace{1cm} (2.52)

where we separated the terms that depend on the stock’s own supply \( u_i \).

2.9.2 Active Manager

The end investor’s utility depends only on the Sharpe ratio of the market and the appraisal ratio of the active portfolio. Hence, the active manager’s problem is to maximize the end investor’s appraisal ratio, subject to the condition that the manager breaks even:

$$\max_{\{\{u_j\} \in \mathcal{S}\}} \frac{\sum_{j=1}^{N_S} v_j \alpha_j - f \sum_{j=1}^{N_S} |v_j|}{\sqrt{\sum_{j=1}^{N_S} v_j^2 \sigma_j^2}}$$

s.t. \( f \sum_{j=1}^{N_S} v_j \alpha_j - f \sum_{j=1}^{N_S} |v_j| \geq C. \) \hspace{1cm} (2.53)

We write the Lagrangian of this problem as

$$\frac{\sum_{j=1}^{N_S} v_j \alpha_j - f \sum_{j=1}^{N_S} |v_j|}{\sqrt{\sum_{j=1}^{N_S} v_j^2 \sigma_j^2}} + \lambda \left[ f \sum_{j=1}^{N_S} v_j \alpha_j - f \sum_{j=1}^{N_S} |v_j| \right] \geq C. \hspace{1cm} (2.54)$$
The first-order conditions with respect to \( v_i \) and \( f \) yield:

\[
\frac{\alpha_i - f(s(v_i))\sqrt{\Sigma_{j=1}^{N^S} v_j\sigma_j^2} - \Sigma_{j=1}^{N^S} v_j\alpha_j - \frac{1}{2} \sum_{j=1}^{N^S} \left( \frac{\Sigma_{j=1}^{N^S} v_j\alpha_j - f(s(v_j))}{\Sigma_{j=1}^{N^S} v_j\sigma_j^2} \right) s(v_i) v_i}{\Sigma_{j=1}^{N^S} v_j^2 \sigma_j^2} + \frac{\lambda}{\Sigma_{j=1}^{N^S} v_j^2 \sigma_j^2} \left( \frac{\sum_{j=1}^{N^S} v_j \alpha_j - f(s(v_j))}{\Sigma_{j=1}^{N^S} v_j \sigma_j^2} \right) s(v_i) v_i = 0
\]

where \( s(v_i) \) indicates the sign of \( v_i \). We take the term containing \( \lambda \) from the latter first-order condition and substitute it into the former, which allows us to get rid of the sign functions. This gives us the portfolio weights:

\[
v_i = \frac{\alpha_i}{\sigma_i^2} \left( 1 + \frac{\lambda f}{\gamma} \right) \left( \frac{1}{\sum_{j=1}^{N^S} v_j \sigma_j^2} \right) \left( \sum_{j=1}^{N^S} v_j \sigma_j^2 \right) \left( \frac{1}{\sqrt{\sum_{j=1}^{N^S} v_j^2 \sigma_j^2} + 2 \frac{\lambda f}{\gamma} \sum_{j=1}^{N^S} v_j |\sigma_j|} \right).
\]

(2.55)

The weights are thus proportional to \( \frac{\alpha_i}{\sigma_i^2} \). We normalize the portfolio by requiring that \( \sum_{v_i > 0} v_i = 1 \), which produces the final expression for the portfolio weights:

\[
v_i = \frac{1}{\sum_{\alpha_i > 0} \frac{\alpha_i}{\sigma_i^2}} \cdot \frac{\alpha_i}{\sigma_i^2}.
\]

(2.56)

### 2.9.3 Analogy with a CARA Investor

Modeling the active manager as a CARA investor with a coefficient of absolute risk aversion \( \gamma \), we let him solve the following maximization problem:

\[
\max_{\{W_i\}} E \left[ -\exp \left( -\gamma \overline{W}_{a,1} \right) \right]
\]

s.t. \( \overline{W}_{a,1} = W_a + \sum_{i=1}^{N^S} W_i \tilde{R}_i \)

(2.57)

where \( \tilde{R}_i = \alpha_i + \beta_i \tilde{R}_m + \tilde{e}_i \) is the excess return on stock \( i \). Without loss of generality, we construct \( N^S \) uncorrelated hybrid securities \( \tilde{z}_i \) where each such security consists of one unit of a stock and a market hedge. The payoff of security \( i \) will then be \( \tilde{z}_i = \alpha_i + \tilde{e}_i \), and its price today will be \( P_{\tilde{z}_i} = P_i + b \eta \). We then express the active manager’s problem as:

\[
\max_{\{W_i\}} E \left[ -\exp \left( -\gamma \overline{W}_{a,1} \right) \right]
\]

s.t. \( \overline{W}_{a,1} = W_a + \sum_{i=1}^{N^S} W_i \tilde{R}_i + W_m \tilde{R}_m \)

(2.58)

where

\[
\tilde{R}_{\tilde{z}_i} = \frac{\tilde{z}_i}{P_{\tilde{z}_i}} - 1 = \frac{\alpha_i + \tilde{e}_i}{P_{\tilde{z}_i}} - 1 = \frac{\alpha_i + \tilde{e}_i - (P_i + b \eta)}{P_i + b \eta}
\]

(2.59)
is the excess return on hybrid security \( i \). Since the payoffs of the hybrid securities are independent, the CARA investor will have a dollar demand for security \( i \) of

\[
W_i = \frac{E[\tilde{R}_i]}{\gamma \text{Var}[\tilde{R}_i]} = \frac{1}{\gamma} \left[ \frac{(a_i - (P_i + b_i \eta))(P_i + b_i \eta)}{\sigma_{\tilde{e}_i}^2} \right] = \left( \frac{a_i - b_i \eta - P_i}{\gamma \sigma_{\tilde{e}_i}^2} \right) \frac{P_i}{P_i + b_i \eta}.
\]  

(2.60)

As each hybrid security \( i \) consists of one share of stock \( i \), the implied dollar demand for stock \( i \) is then

\[
W_i = W_i \frac{P_i}{P_i} = \frac{a_i - b_i \eta - P_i}{\gamma \sigma_{\tilde{e}_i}^2} \frac{P_i}{P_i + b_i \eta} = \frac{(a_i - b_i \eta - P_i)P_i}{\gamma \sigma_{\tilde{e}_i}^2}.
\]  

(2.61)

To obtain a more intuitive expression, we substitute in the abnormal return of the stock \( \alpha_i \) and the idiosyncratic variance of stock return \( \sigma_i^2 \) (not to be confused with payoff variance \( \sigma_{\tilde{e}_i}^2 \)):

\[
W_i = \frac{1}{\gamma} \left[ \frac{a_i - b_i \eta - P_i}{P_i} \right] \frac{P_i^2}{\sigma_i^2} = \frac{\alpha_i}{\gamma \sigma_i^2}.
\]  

(2.62)

Note that each position in a hybrid security \( i \) will also generate a dollar demand of \( b_i \) for the market portfolio (to hedge market risk) and a dollar demand of \(-b_i (1 + \eta)\) for the risk-free asset. In equilibrium it will turn out that these hedging demands from the long and short positions perfectly cancel out as the active manager holds symmetric share positions around zero, so we do not need to address the question of whether the active manager should hedge market risk of the stock positions on his own or leave it to the end investor.

An unconstrained CARA investor would also have a "speculative" dollar demand of

\[
W_m = \frac{\eta}{\gamma \sigma_m^2}
\]  

(2.63)

for the market portfolio directly. We set this demand equal to zero because the end investor should not reward the active manager for investing in the market portfolio. In the previous optimization problem (2.53) of the manager we did the same thing implicitly as we considered only abnormal returns and the market portfolio of course has an abnormal return of zero.

### 2.9.4 Equilibrium

We denote the supply of stock \( i \) left to the active manager as \( u_i \). For the market to clear, the dollar supply has to equal the dollar demand, and this gives us the stock price:

\[
\begin{align*}
\alpha_i P_i & = \frac{\alpha_i}{\gamma \sigma_i^2} = \frac{(a_i - b_i \eta - P_i)P_i}{\gamma \sigma_{\tilde{e}_i}^2} \\
P_i & = \frac{a_i - b_i \eta - \gamma \sigma_{\tilde{e}_i}^2 u_i}{P_i}
\end{align*}
\]  

(2.64)

(2.65)

The alpha of the stock is then:

\[
\alpha_i = \frac{E[\tilde{x}_i]}{P_i} - \beta_i \eta - 1 = \frac{a_i - b_i \eta - P_i}{P_i} = \frac{\gamma \sigma_{\tilde{e}_i}^2 u_i}{P_i}.
\]  

(2.66)
The market portfolio has an alpha of zero by construction. Hence,

\[
\alpha_m = \frac{\sum_{i=1}^{N_S} P_i \alpha_i}{\sum_{i=1}^{N_S} P_i} = 0
\]  
\[
\sum_{i=1}^{N_S} P_i \alpha_i = \sum_{i=1}^{N_S} P_i \gamma \sigma_{\epsilon, u_i}^2 = \gamma \sigma_\epsilon^2 \sum_{i=1}^{N_S} u_i = 0
\]  
\[
\sum_{i=1}^{N_S} u_i = 0.
\]  

Since \( u_p + u_m + u_t = 1 \), and since the passive manager's position \( u_p \) is constant across stocks while the noise trader's position \( u_m \) is distributed as \( N \left( 0, \sigma_u^2 \right) \), the above equation implies the same distribution for the active manager's equilibrium share holdings \( u_t \):

\[
u_i \sim N \left( 0, \sigma_u^2 \right).
\]  

As the noise trader and the active manager hold an average of zero of each stock, the passive manager has to hold the entire supply of 1 share, and hence he will hold the entire market portfolio:

\[
u_p = \frac{W_p}{P_m} = 1.
\]  

Denoting the price of the market portfolio as \( P_m \), its expected payoff as \( a_m \), and the dollar variance of that payoff as \( \sigma_m^2 \), and plugging in the end investor's allocation to the passive manager, we obtain the equilibrium market risk premium

\[
1 = \frac{W_p}{P_m} = \left( \frac{\eta}{\gamma e \sigma_m^2} \right) \frac{1}{P_m} = \left( \frac{\eta P_m^2}{\gamma e \sigma_M^2} \right) \frac{1}{P_m} = \frac{\eta a_m}{\gamma e \sigma_M^2 (1 + \eta)}
\]  
\[
\eta = \frac{\gamma e \sigma_M^2}{a_m - \gamma e \sigma_M^2}
\]  

and the equilibrium allocation to the passive manager

\[
W_p = P_m = \frac{a_m}{1 + \eta} = a_m - \gamma e \sigma_M^2.
\]  

To find out the allocation to the active manager, we need to find the before-fees alpha of the manager:

\[
\alpha_{bf} = \frac{\sum_{i=1}^{N_S} P_i u_i \alpha_i}{\sum_{u_i > 0} P_i u_i}.
\]  

The cost of the portfolio is determined by the long positions, so only the long positions show up in the denominator. The numerator can be expressed as

\[
\sum_{i=1}^{N_S} P_i u_i \alpha_i = \sum_{i=1}^{N_S} \gamma \sigma_{\epsilon, u_i}^2 u_i^2 = \gamma \sigma_\epsilon^2 \sum_{i=1}^{N_S} u_i^2 = \gamma \sigma_\epsilon^2 N_S \sigma_u^2.
\]  

For this aggregation, we used the assumption that there is a continuum of stocks with a measure of \( N_S \), so \( \frac{1}{N_S} \sum_{i=1}^{N_S} u_i^2 = E \left[ u_i^2 \right] = \sigma_u^2 \). If we do not make the assumption, our results will be in the terms of particular
realizations of all the \( u_i \)'s \((N_S\ of\ them)\), so the increase in mathematical rigor would come at the high cost of eliminating the simplicity and transparency of the equilibrium expressions. Due to the law of large numbers, this approximation does not affect our results in any meaningful way. Similarly for the denominator, we get

\[
\sum_{u_i > 0} P_i u_i = \sum_{u_i > 0} (a_i - b_i \eta - \gamma \sigma_i^2 u_i) u_i = (a - b) \sum_{u_i > 0} u_i - \gamma \sigma_u^2 \sum_{u_i > 0} u_i
\]

\[
= \frac{N_S \sigma_u}{2} \left[ \sqrt{\frac{2}{\pi}} (a - b) - \gamma \sigma_u^2 \sigma_u \right]
\]

(2.77)

We also need the idiosyncratic variance of the active manager's portfolio. That is simply

\[
\sigma^2_a = \frac{\sum_{i=1}^{N_S} u_i^2 \sigma_i^2}{\left( \sum_{u_i > 0} P_i u_i \right)^2} = \frac{4 \sigma^2}{N_S \left[ \sqrt{\frac{2}{\pi}} (a - b) - \gamma \sigma_u^2 \sigma_u \right]^2}.
\]

(2.78)

The fee of the active manager as a percentage of the cost of the portfolio is given by

\[
f_a = \frac{\sum_{i=1}^{N_S} |W_i| P_i u_i}{\sum_{u_i > 0} P_i u_i}
\]

(2.79)

where the numerator is

\[
\sum_{i=1}^{N_S} |W_i| = \sum_{i=1}^{N_S} P_i |u_i| = - \sum_{u_i < 0} P_i u_i + \sum_{u_i > 0} P_i u_i = \sqrt{\frac{2}{\pi}} (a - b) N_S \sigma_u,
\]

(2.80)

and thus we get

\[
f_a = \frac{2 \sqrt{\frac{2}{\pi}} (a - b) f}{\sqrt{\frac{2}{\pi}} (a - b) - \gamma \sigma_u^2 \sigma_u}.
\]

(2.81)

Finally, we can obtain the end investor's allocation to the active manager which depends on the after-fees alpha:

\[
W_a = \frac{\alpha_{af} - f_a}{\gamma_c \sigma_a^2} = \frac{N_S \left[-\sqrt{\frac{2}{\pi}} (a - b) f + \gamma \sigma_u^2 \sigma_u \right]}{2 \gamma_c \sigma_u^2} \left[ \sqrt{\frac{2}{\pi}} (a - b) - \gamma \sigma_u^2 \sigma_u \right].
\]

(2.82)

Equating this with the cost of the manager's portfolio \((2.77)\), we obtain the simple formula for the effective risk aversion of the manager:

\[
\gamma = \gamma_c + \sqrt{\frac{2}{\pi}} \left( \frac{a - b \eta}{\sigma_u^2 \sigma_u} \right) f.
\]

(2.83)

The value of the proportional fee \( f \) allows the manager to exactly cover his fixed dollar cost \( C \):

\[
C = f \sum_{i=1}^{N_S} |W_i| = \sqrt{\frac{2}{\pi}} (a - b) N_S \sigma_u f
\]

(2.84)

\[
f = \frac{C}{\sqrt{\frac{2}{\pi}} (a - b) N_S \sigma_u}.
\]

(2.85)
2.10 Figures

Figure 2-1: July 2002 deletion of non-U.S. firms from S&P 500.
Figure 2-2: The basic setup for the model.
Chapter 3

Selection of an Optimal Index Rule for an Index Fund

3.1 Introduction

Index funds are a peculiar investment vehicle. Their investment strategy consists of picking a stock market index such as the S&P 500 and then mechanically buying the stocks in the index in an attempt to replicate the index returns as closely as possible. While active fund managers prefer not to disclose which stocks they buy and sell (until after the fact), index funds explicitly declare their strategy to all market participants. The behavior of the index funds is therefore completely predictable to the other market participants, at least to the extent that the composition of the stock index itself is predictable.

This kind of predictable behavior seems to bring a cost with it. When a new stock is added to the index, the price of this stock generally goes up. Index funds will then buy the stock at the new price, i.e. after the price increase has already occurred. When a stock is deleted from the index, the price goes down, and now the index funds have to sell the stock right after its price has dropped. As a result, the index funds systematically end up buying high and selling low. This seems like a bad deal for them, even when it does not increase their tracking error.

The price premium for S&P 500 index stocks was about 15% of the stock price in 2000 (see Chapter 1). That year, about 12% of the index stocks were deleted from the index and replaced with new stocks. The incoming stocks accounted for 5.6% of the value of the index,
while the outgoing stocks accounted for 0.4%. S&P 500 index funds had to update their portfolios accordingly, buying added stocks with the index premium and selling deleted stocks without the index premium. This stands in contrast to an index-neutral strategy which would also hold some non-index stocks and would therefore gain the index premium on some stocks that are added to the index. S&P 500 index stocks constitute around 80% of the market value of large and medium-cap firms, so an index-neutral strategy would hold about this proportion of the portfolio in index stocks. The difference in portfolio returns between the index and a similar index-neutral portfolio represents what we label an index turnover cost, and we estimate it to be around 0.7% per year for the S&P 500.

Given that some index funds charge annual management fees of 0.18% of assets under management, an additional certain and recurring (while somewhat hidden) annual cost of 0.7% seems rather high. For an index investor with a 30-year horizon, this would cut his final wealth by 19%. The absolute magnitude of the cost is also substantial: at the end of 2000, the Vanguard 500 Index Fund alone had about $100 billion in assets, which implies an annual index turnover cost of $0.7 billion to the investors in that single fund. Furthermore, all these numbers assume that the price impacts of index addition and deletion are permanent – if instead they fully reverse over time, the index funds will have to pay a separate cost when buying and selling, so the total costs will be even higher.

So far, academic literature has focused on the market-efficiency implications of index addition and on the trading strategies that would exploit the index effect. Yet next to nothing has been said about the other side of the deal: the index funds that would be taken advantage of by index arbitrageurs playing the “S&P game.” While the question for the arbitrageurs is how to maximize their gains from the index effect, the question for the index funds is how to minimize their costs due to the index effect.

The index turnover cost arises from following the mechanical investment strategy of buying the stocks in the index. But going a little further, perhaps it is not only being in the index but also the way the index stocks are selected that determines the index premium. A priori we can think of several ways how the index selection rule would affect the price dynamics and ultimately the performance of the index funds. The index selection rule could be based on e.g. market capitalization, or some exogenous random variable, or a deterministic rule, or a
combination of these. For example, the Russell indices such as the Russell 1000 and Russell 2000 are mechanically based on market capitalization, while the S&P indices such as the S&P 500 and S&P SmallCap 600 are also influenced by other partially unobservable (random) criteria. Different rules can result in great differences in the frequency of updating the index, and thus in the frequency at which the index funds have to pay the cost due to the index premium. The rules may also have effects on the size of the index premium and hence on the size of the cost paid each time the index is updated. What are the effects of the different rules on price dynamics and on the performance of the index funds? What would be the best rule for the index funds, i.e. the one that minimizes the expected value of the index turnover cost?

At this point we might also wonder whether the index funds should care about the index turnover cost in the first place. After all, do they not exist for the sole purpose of tracking an index? As long as the index itself suffers the same cost, why would it matter? The answer is that the index funds already care about their expected return and not only about their sensitivity to the index. The primary concern of a typical index fund is minimizing both its tracking error and its transaction costs due to fund inflows and outflows. They choose not to track the index perfectly because then they would incur transaction costs that would reduce their expected returns too much. Thus it seems natural that if the index fund investors care about lower expected returns due to transaction costs, they should also care about lower expected returns due to the index turnover cost.

Unfortunately for the index investors, this cost is hard to detect empirically using realized returns. If in the second half of the 1990s the index premium has reduced the expected annual return on the S&P 500 index by an average of 0.3%-0.4% relative to the expected return on an index-neutral strategy (note that the index premium and the index turnover cost have been growing all this time), any such shift in the mean is easily swamped by the 16% annual volatility of the index. The large influx of money into S&P 500 index funds during this time has also produced capital gains on the existing stocks in the index, increasing the realized index returns and thus concealing the lower expected future index returns. These may be reasons why the index turnover cost seems to have largely eluded the attention of investors and investment advisors who keep touting the low management fees charged by index funds.¹

¹The first time we noticed some of these concerns raised publicly was in the May 2002 issue of Institutional
In this chapter we build a model to investigate the behavior of the index premium and the performance of the index funds for different index selection rules. To keep our analysis simple and manageable, we focus exclusively on the index premium that would arise in the presence of rational investors and frictionless markets. For us the primary interesting phenomenon is the index turnover cost which arises from the fundamental nature of index funds and which cannot be avoided even in a perfectly functioning frictionless market.

We find that the rational anticipation of the future index composition reflected in stock prices today eliminates any first-order differences in index fund performance across index selection rules. If the probability or frequency of a future index change is high, this results in a lower index premium, and the two effects cancel out so that the expected index turnover cost is not affected. However, as the index funds become large, the other investors become less diversified, and this induces hedging motives which do create differences in index turnover cost across the index rules. In particular, a market-cap rule (buy large stocks) represents a positive-feedback strategy for the index funds which amplifies the fundamental volatility of stocks and increases the expected index turnover cost.

A more elaborate model would incorporate transaction costs, institutional rigidities, and perhaps even some investor irrationality. Introducing some specific form of transaction costs would be a conceptually straightforward extension of the current model, although we might then have to resort to numerical analysis rather than analytical expressions. If we judge that some predictable investor irrationality also affects the index premium, the real challenge is to establish the exact nature of that deviation from rationality so that we can investigate its impact on the index premium under different index selection rules. Even if we are able to discover this irrationality with enough confidence to include it in a model, perhaps by that time the actual market participants will have figured out the same irrationality and it will have been arbitraged away already. These issues would have to be addressed in future empirical work.

The chapter proceeds as follows. Section 2 briefly discusses index funds in the context of the index turnover cost, while most of the relevant background information was provided in Chapter 1. Section 3 presents our basic discrete-time model. Section 4 analyzes the equilibrium of the model for different index selection rules. Section 5 extends the model to continuous

Investor ("Is Time Running Out for the S&P 500?").
time. Section 6 discusses some interpretations of the model, and section 7 concludes. All mathematical details are in the appendix, which also contains the graphs for the comparative statics results.

3.2 Behavior of Index Funds

Chapter 1 discusses the background for this chapter, including the basics of the S&P 500 and Russell 2000 as well as their associated price premia and turnover costs.

For most index funds today, the first question would probably be whether the index turnover cost increases their tracking error. Index additions and deletions for both the S&P and Russell indices are announced several trading days before they become effective, so prices move before the indices are updated. As a result, the exact same index turnover cost suffered by the index funds is also suffered by the index itself, and thus it does not contribute to the tracking error of an index fund.

According to Frank Russell Co., their index selection and updating policy is the result of trade-off: If they updated the indices more often than once a year, the annual turnover of the indices would increase which, in turn, would generate more transaction costs to index funds tracking the Russell indices. But if they updated the indices less frequently, the indices would not remain representative of the market segments for which they are supposed to proxy. Russell judges annual updating to strike a good balance between the two objectives. Hence, the costs of the indexers do influence Russell’s choice of an index rule, although this analysis appears to focus exclusively on transaction costs while not dealing with the index turnover cost (which arises entirely due to the index premium).

Nevertheless, the index turnover cost to the index investors is real, although it is hidden by the usual volatility of the stock market and the recent index fund inflows. It seems like most index investors are either unaware of this additional cost or they think paying it is still better than paying the high fees of active fund managers. There has been a lot of research recently on one side of the issue (i.e. how to take advantage of index funds), so perhaps it is just a matter of time when the other side of the issue (i.e. the index turnover cost) also receives some attention.
3.3 The Model

3.3.1 Motivation

We need to build a model to be able to investigate the price dynamics in the presence of an index fund. Once we get the price dynamics, we can measure the index turnover cost and consequently the performance of the index fund. Then we can change the index rule and see what effect this has on the performance of the index fund.

The model is intended to be as simple as possible while still capturing the index premium and the dynamics associated with rational anticipation of a possible index revision in the future. Later we confirm our results in a more general continuous-time setup.

3.3.2 Model Setup

We construct a dynamic two-period model (with three dates) which consists of the following:

- **Assets:**
  
  - 2 risky stocks, each having a random payoff $\pi_{i1} + \pi_{i2} + \pi_{i3}$ at the terminal date and existing in positive net supply. Let $\pi_{ij} \sim N(\mu, \sigma^2)$. The payoffs $\pi_{ij}$ are iid across time (labeled $j$) but contemporaneously across stocks (labeled $i$) they have a correlation $\rho$. The supply of each stock is 1.
  
  - A risk-free asset with interest rate $r = 0$ and existing in infinitely elastic supply.

- **Index:**

  - Under the market-cap rule, the index consists of 1 risky stock with the highest market capitalization (stock price). We also allow for other index selection rules.

- **Agents:**

  - An index fund. Each period the index fund wants to hold a fraction $h$ of the supply of the stock that is in the index that period.
A representative investor with constant absolute risk aversion and utility over terminal wealth. Its objective function is

$$
\max E[-\exp(-bW_{t3})]
$$

(3.1)

where $W_{t3}$ represents the investor’s terminal wealth and $b$ is the coefficient of absolute risk aversion. We will usually refer to the representative investor as the “CARA investor(s)” or simply “investor(s),” as opposed to the “index fund.”

• Timeline:

- $t = 1$: News of the first dividend shock $\pi_{i1}$ revealed for each stock $i$. (The actual distribution of the dividend $\pi_{i1}$ will take place later at $t = 3$.) The index fund and the investors trade. Prices are formed. The composition of the index is determined.

- $t = 2$: News of the second dividend shock $\pi_{i2}$ revealed for each stock $i$. The index fund and the investors trade. Prices are formed. The composition of the index is determined again, and the index is possibly revised.

- $t = 3$: News of the third dividend shock $\pi_{i3}$ revealed for each stock $i$. The risky assets pay their terminal dividends.

To see how the index revision and its anticipation will influence the market participants and prices, we need to determine prices at three different points in time which in turn determines two consecutive stock returns (including two returns for the stock index, the composition of which may change between the two periods).

We choose CARA utility for the representative investor because it is analytically convenient as it eliminates any wealth effects across the two stocks induced by the representative investor. The wealth effects are not our primary interest here, and they would merely blur the interesting effects arising from the presence of the index fund.

We also assume the index fund buys a fraction $h$ of the supply of the index stock, rather than simply investing all its wealth in the index stock. This is to eliminate any wealth effects induced by the index fund. This trading strategy is simply a buy-and-hold for the index stock, but if the index is updated, the strategy may not be self-financing. Yet even in those cases.
the strategy is still very close to being self-financing, so this assumption should not change our results in an economically meaningful way, but it does make the analytical expressions much more tractable.

The model is frictionless, so transaction costs will be zero and the index fund will have zero tracking error. This allows us to focus exclusively on the index premium which is not related to the tracking error of the index fund but does result in the index itself having lower returns.

3.3.3 Equilibrium

In equilibrium, we require that the index fund is fully invested in the index stock at each time $t$, the representative CARA investor holds his optimum portfolio, and the stock market clears.

The model should be solved backwards, starting at the terminal date. The situation at $t = 3$ is clear: the investors get their payoffs from their asset holdings and their utilities are realized. We start by solving for the equilibrium at $t = 2$, i.e. we find the market-clearing prices for the two stocks at that time. Once this is done, we can solve the equilibrium at $t = 1$.

Equilibrium at $t = 2$

The investor's objective function at $t = 2$ is:

$$\max_{\alpha_{12}, \alpha_{22}} E_2[-\exp(-bW_{t3})]$$

subject to:

$$W_{t3} = W_{t2} + \alpha_{12}(p_{13} - p_{12}) + \alpha_{22}(p_{23} - p_{22})$$

(3.2)

Here $E_2[.]$ denotes the expectation operator given the information available at $t = 2$, $W_{t3}$ is the wealth of the investor at $t = 3$, $b$ is the investor's coefficient of absolute risk aversion, $\alpha_{12}$ is the investor's holding of stock 1 at $t = 2$ (i.e. from $t = 2$ to $t = 3$), and $p_{13}$ is the price of stock 1 at $t = 3$. This is a standard CARA-normal setup which is easy to solve.

Let $\{1\}$ represent the event that stock 1 is in the index at $t = 2$, and similarly for stock 2. At any time exactly one of the stocks will be in the index, so the events $\{1\}$ and $\{2\}$ are exhaustive and mutually exclusive. $1_{\{1\}}$ is the indicator function and it is equal to 1 if $\{1\}$ has happened and 0 otherwise.
The equilibrium stock holdings and stock prices are the following, where $W_{f2}$ is the wealth of the index fund and $a_{f12}$ and $a_{f22}$ are its stock holdings at $t = 2$:

\[ a_{f12} = 1_{\{1\}} \frac{W_{f2}}{p_{12}} \]  
\[ a_{f22} = 1_{\{2\}} \frac{W_{f2}}{p_{22}} \]  
\[ a_{i12} = 1 - 1_{\{1\}} \frac{W_{f2}}{p_{12}} \]  
\[ a_{i22} = 1 - 1_{\{2\}} \frac{W_{f2}}{p_{22}} \]  
\[ p_{12} = \frac{1}{2} \left[ E_2[p_{13}] - (1 + \rho) b \sigma^2 + \sqrt{(E_2[p_{13}] - (1 + \rho) b \sigma^2)^2 + 4b \sigma^2 W_{f2} (\rho + 1_{\{1\}}(1 - \rho))} \right] \]  
\[ p_{22} = \frac{1}{2} \left[ E_2[p_{23}] - (1 + \rho) b \sigma^2 + \sqrt{(E_2[p_{23}] - (1 + \rho) b \sigma^2)^2 + 4b \sigma^2 W_{f2} (\rho + 1_{\{2\}}(1 - \rho))} \right] \]

As we go recursively back in time, these expressions become unmanageable very quickly. We therefore make a slight simplification by assuming that the index fund will always hold a fraction $\mu$ of the index stock. This produces a much simpler linear equilibrium:

\[ a_{f12} = 1_{\{1\}} \mu \]  
\[ a_{f22} = 1_{\{2\}} \mu \]  
\[ a_{i12} = 1 - 1_{\{1\}} \mu \]  
\[ a_{i22} = 1 - 1_{\{2\}} \mu \]  
\[ p_{12} = E_2[p_{13}] - (1 + \rho) b \sigma^2 + \rho \mu b \sigma^2 + 1_{\{1\}}(1 - \mu) h b \sigma^2 \]  
\[ = \pi_{11} + \pi_{12} + \mu - (1 + \rho) b \sigma^2 + \rho \mu b \sigma^2 + 1_{\{1\}}(1 - \mu) h b \sigma^2 \]  
\[ p_{22} = E_2[p_{23}] - (1 + \rho) b \sigma^2 + \rho \mu b \sigma^2 + 1_{\{2\}}(1 - \mu) h b \sigma^2 \]  
\[ = \pi_{21} + \pi_{22} + \mu - (1 + \rho) b \sigma^2 + \rho \mu b \sigma^2 + 1_{\{2\}}(1 - \mu) h b \sigma^2 \]

Perhaps the best way to get the intuition about the stock prices is to assume that the dividend shocks across the two stocks are uncorrelated ($\rho = 0$). We then have the following
prices:

\[ p_{12} = E_2 [p_{13}] - b\sigma^2 + 1_{\{1\}} h b \sigma^2 \]  
\[ p_{22} = E_2 [p_{23}] - b\sigma^2 + 1_{\{2\}} h b \sigma^2 \]  

(3.18)  

(3.19)

Each stock will be worth the expected payoff next period minus the risk premium, plus a premium for being in the index now. There is no additional discounting of payoffs because the risk-free rate is zero. The index premium is positive because the index fund is bearing some of the risk of the index stock, thus effectively reducing the supply of the index stock and reducing the necessary risk premium in equilibrium.

If the two stocks have correlated fundamentals \((\rho > 0)\), the risk premium for holding each stock will increase to \((1 + \rho) b \sigma^2\). Now both stocks will also share a fraction \(\rho\) of the index premium, while the remaining index premium \((1 - \rho) h b \sigma^2\) will show up in the price of the index stock alone.

**Equilibrium at \(t = 1\)**

At \(t = 1\) the objective function of the investor is:

\[
\max_{a_{i11}, a_{i21}, a_{i12}, a_{i22}} E_1 \left[ -\exp \left( -b W_{i3} \right) \right] 
\]

subject to:

\[ W_{i3} = W_{i2} + a_{i12} (p_{13} - p_{12}) + a_{i22} (p_{23} - p_{22}) \]  
\[ W_{i2} = W_{i1} + a_{i11} (p_{12} - p_{11}) + a_{i21} (p_{22} - p_{21}) \]  

(3.21)  

(3.22)

Because the investment opportunities at \(t = 2\) turn out to be independent of the dividend shocks realized between \(t = 1\) and \(t = 2\), we can simplify the objective function to a one-period-ahead maximization (see the Appendix):

\[
\max_{a_{i11}, a_{i21}} E_1 \left[ -\exp \left( -b W_{i2} \right) \right] 
\]

subject to:

\[ W_{i2} = W_{i1} + a_{i11} (p_{12} - p_{11}) + a_{i21} (p_{22} - p_{21}) \]  

(3.23)  

(3.24)
The problem here is that given the information we have at \( t = 1 \), the investor's wealth \( W_{12} \) is not normally distributed, so we no longer have a convenient CARA-normal setup. Although the dividend shocks which determine \( p_{12} \) and \( p_{22} \) are normally distributed, these shocks can also determine the composition of the index, and this introduces a discreteness into the distribution of prices at \( t = 2 \), resulting in a non-normal distribution of prices and the wealth \( W_{12} \).

However, we can still calculate the equilibrium prices in closed form. As always in the absence of arbitrage, we can write the price of stock \( i \) at \( t = 1 \) as a function of the state-price density (or the stochastic discount factor) \( m_2 \) at \( t = 2 \):

\[
p_{i1} = E_1 [p_{i2} m_2].
\] (3.25)

In this case it turns out that we can solve for \( m_2 \) in closed form and thus calculate the stock prices at \( t = 1 \) (see the Appendix). We get:

\[
m_2 = \frac{\exp \left( -ba_{i11} \left( \pi_{12} + 1 \{1 \} (1 - \rho) h \sigma^2 \right) - ba_{i21} \left( \pi_{22} + 1 \{2 \} (1 - \rho) h \sigma^2 \right) \right)}{E_1 \left[ \exp \left( -ba_{i11} \left( \pi_{12} + 1 \{1 \} (1 - \rho) h \sigma^2 \right) - ba_{i21} \left( \pi_{22} + 1 \{2 \} (1 - \rho) h \sigma^2 \right) \right]}. \] (3.26)

In equilibrium, \( a_{i11} = 1 - h \) and \( a_{i21} = 1 \) if stock 1 is in the index at \( t = 1 \), and \( a_{i11} = 1 \) and \( a_{i21} = 1 - h \) otherwise, so these are known values at \( t = 1 \). If we plug in the equilibrium prices at \( t = 2 \) of stocks 1 and 2, we get the following equilibrium prices at \( t = 1 \):

\[
p_{11} = \pi_{11} + \mu - (1 + \rho) b \sigma^2 + \rho h \sigma^2 + E_1 \left[ \left( \pi_{12} + 1 \{1 \} (1 - \rho) h \sigma^2 \right) m_2 \right] \] (3.27)

\[
p_{21} = \pi_{21} + \mu - (1 + \rho) b \sigma^2 + \rho h \sigma^2 + E_1 \left[ \left( \pi_{22} + 1 \{2 \} (1 - \rho) h \sigma^2 \right) m_2 \right]. \] (3.28)

Based on these expressions, it is clear that there will be two sources of risk that are priced: the fundamental risk of \( \pi_{42} \) (i.e. the dividend news next period), and the risk associated with index uncertainty. In the presence of an index fund (\( h > 0 \)), the index selection brings with it a positive index premium to the stock that happens to be in the index next period. A higher correlation across the fundamentals of the stocks reduces this effect but does not eliminate it.

The expectations depend on how the index is selected at \( t = 2 \). If the index selection is independent of the fundamentals of the stocks, we can compute the expectations easily. If it is not, such as if the index selection is based on market capitalization, we can still calculate
the expectations analytically, but the calculations are much trickier and the final results look less intuitive. The complete analytical expressions for the stock prices (after the expectations have been worked out) will be discussed in the next section in the context of the different index selection rules.

3.4 Analysis of Equilibrium

We now turn to an analysis of the equilibrium. We are interested the comparative statics of the model at $t = 1$, so we examine prices at $t = 1$ and returns from $t = 1$ to $t = 2$ to see how they change as a function of the parameters of the model.

First, we want to see if there is anything interesting in the price dynamics due to the presence of the index fund. In particular, how does the anticipation of a possible future change in the index affect prices and returns now? How does this vary across different index selection rules?

Second, how does the index selection rule affect the performance of the index fund? Should the index be selected endogenously, e.g. based on market cap? Or should it be selected exogenously, yet perhaps still randomly?

To keep the analysis as intuitive as possible, we set the correlation across the fundamental news to zero ($\rho = 0$). This makes some of the interesting effects more visible. As evident from equations (3.26), (3.27), and (3.28), increasing the parameter $\rho$ from zero will only gradually reduce the effects due to index selection – qualitatively the effects will be the same (unless of course we set $\rho = 1$ in which case the two stocks effectively collapse into one stock).

3.4.1 Market-Cap Rule

Let us investigate a situation where the index is selected each period based on market capitalization, and the stock with the higher market capitalization will be chosen for the index. Both stocks have the same supply, so the stock with the higher price will be in the index.

In fact there is a region where the index selection is somewhat indeterminate: if the fundamentals and therefore the prices of the two stocks are close to each other, then the additional buying by the index fund can make either stock price higher than the other, and thus either stock could be in the index. To resolve this indeterminacy we choose the index stock to be the
one that would have a higher market capitalization without the index fund. In the context of this model it is the stock that has received better dividend shocks ($\pi_{11}$ at $t = 1$ or the sum of dividend shocks $\pi_{i1} + \pi_{i2}$ at $t = 2$).

The expectations in the expressions (3.26), (3.27), and (3.28) can be evaluated analytically now that we have fixed an index selection rule (see the Appendix). Let

$$A = \frac{\pi_{21} - \pi_{11}}{\sigma \sqrt{2(1 - \rho)}} - b \sigma (a_{21} - a_{11}) \sqrt{\frac{1 - \rho}{2}}. \tag{3.29}$$

This represents the threshold value that a standard normal dividend shock to stock 1 has to exceed so that stock 1 will be chosen for the index at $t = 2$. The greater the relative dividend shock to stock 1 at $t = 1$, the less the necessary threshold value at $t = 2$. The stock price is then given by:

$$p_{11} = \pi_{11} + 2\mu - 2(1 + \rho) b \sigma^2 + \frac{3}{2} \rho h b \sigma^2 + \frac{1}{2} h b \sigma^2 + \frac{1}{2} b \sigma^2 (a_{21} - a_{11}) (1 - \rho) \exp \left[-\frac{1}{4} b^2 \sigma^2 h^2 (1 - \rho)\right] + \frac{C}{\exp \left[-a_{11} h b^2 \sigma^2 (1 - \rho)\right] \Phi (A) + \exp \left[-a_{11} h b^2 \sigma^2 (1 - \rho)\right] [1 - \Phi (A)]} \tag{3.30}$$

$$+ \frac{\exp \left[-a_{21} h b^2 \sigma^2 (1 - \rho)\right] \Phi (A) + \exp \left[-a_{11} h b^2 \sigma^2 (1 - \rho)\right] [1 - \Phi (A)]}{\exp \left[-a_{21} h b^2 \sigma^2 (1 - \rho)\right]} \tag{3.31}$$

where $C$ in turn is defined as follows:

$$C = (1 - \rho) \frac{h b \sigma^2}{\exp \left[-a_{11} h b^2 \sigma^2 (1 - \rho)\right]} \frac{[1 - \Phi (A)]}{\sqrt{\frac{1 - \rho}{2}}} \left[\exp \left[-\frac{1}{4} b^2 \sigma^2 h^2 (1 - \rho)\right] \Phi (A) + \left[\exp \left[-a_{11} h b^2 \sigma^2 (1 - \rho)\right] - \exp \left[-a_{21} h b^2 \sigma^2 (1 - \rho)\right]\right] \phi (A). \tag{3.32}$$

Due to the symmetric model setup, the price of stock 2 is given by a similar expression where we only need to replace all references to stock 1 with stock 2 and vice versa.

**Index Fund Size and Index Change Probability**

First we vary the size of the index fund ($h$) and the fundamentals of each stock at $t = 1$ (i.e. the difference between the first dividend shocks $\pi_{11}$ and $\pi_{21}$). We choose the following parameter
values for the analysis:

\[ b = 1 \]
\[ \mu = 1 \]
\[ \sigma = 0.2 \]
\[ \pi_{11} = 1 \]
\[ h \in [0, 1] \]
\[ \pi_{21} \in [0.5, 1.5] \]

If the coefficient of absolute risk aversion \( b \) is equal to 0.5 and there is no index fund, we get a market risk premium of 8% for a standard deviation of 16% for the period from \( t = 0 \) to \( t = 3 \). When there is an index fund, the coefficient for the other investors would have to be higher. We choose \( b = 1 \) as a reasonable starting point.

The mean of the payoff \( \mu \) is normalized to 1. Its standard deviation \( \sigma \) is picked so that the standard deviation of the unconditional return on each stock is roughly 10%. Using these values, the probability of a negative payoff at \( t = 3 \) is negligible (being an 8.7 standard deviation event).

We vary the index fund size \( h \) between 0 and 1, where 1 means that the index fund wants to hold all of the supply of the index stock.

The initial dividend shock \( \pi_{21} \) of stock 2 varies from 0.5 to 1.5. For \( \pi_{21} < 1 \), stock 1 will be in the index at \( t = 1 \) (since we fixed \( \pi_{11} = 1 \) for stock 1), and for \( \pi_{21} \geq 1 \) stock 2 will be in the index \( t = 1 \). With these parameter values, either extreme means that there is a 2% probability of an index change at \( t = 2 \), while a value of 1.0 means that the probability of an index change at \( t = 2 \) is exactly 50%. The index is selected at \( t = 1 \) and again at \( t = 2 \) based on market capitalization at each point in time.

The results are illustrated in surface plots (see the Appendix).

The price of stock 1 looks like what we would expect. If stock 1 is in the index, its price will be pushed up by the index fund. If the initial dividend shock \( \pi_{21} \) of stock 2 is close to 1 (i.e. the value of \( \pi_{11} \)), there is a relatively high probability that the index will be changed at \( t = 2 \), and this blurs the boundary between being in and out of the index.
The price of stock 2 is affected not only by the index fund but also by what we could call the "supply effect" – increasing the initial dividend shock $\pi_{21}$ directly increases the expected payoff at $t = 3$, so it is essentially identical to adding more units of the risk-free asset to the terminal payoff of stock 2, and hence it will have an impact on the price at $t = 1$. Since the interesting economic effects are muddled here by the supply effect of stock 2, we focus on stock 1 in our analysis.

The expected dollar return\(^2\) on stock 1 looks more surprising on a careful examination. As we might have anticipated, there are two distinct regions: if the stock is in the index at $t = 1$, the expected return will be driven down by the index fund, but if the stock is not in the index, the expected return will be relatively constant. But there are two peculiar effects: First, why does the expected return increase as the index fund gets bigger when stock 1 is out of the index? Should it not stay constant or perhaps decrease slightly, since a larger index fund can only push the prices up and decrease the risk premium? Second, how can the expected return on stock 1 be negative, rather than zero, when stock 1 is in the index and the index fund holds all of it? Would the other investors not want to short stock 1 in this case, and the market for the stock would not clear?

The first effect is explained by the variance of the stock return. As the initial dividend shock $\pi_{21}$ gets close to 1 (i.e. the value of $\pi_{11}$), there is a lot of uncertainty about which stock will be in the index next period. Since the index stock will get the index premium next period, this means there will be a lot of uncertainty about which stock will have the index premium, and this uncertainty creates a new source of risk for the investors holding either stock. A larger index fund will have a larger price impact, thus increasing this risk even more and increasing the expected return to compensate the investors for this risk.

The second effect can be explained by the cross-correlation of returns on the two stocks. If the index fund does not exist ($h = 0$), the cross-correlation is zero. This follows from the independence of the stocks' dividend shocks and the CARA preferences which eliminate any wealth effects across the stocks. As the index fund grows in size, the correlation across the two stocks becomes negative. This is because the two stocks are essentially competing to be in the

\(^2\)Note that whenever we refer to returns, they are dollar returns, not rates of return. With CARA utility, a certain dollar risk exposure tends to produce the same dollar risk premium, while the rate of return would much more complicated to interpret.
index next period: a good dividend shock for stock 1 increases the chances of stock 1 winning and stock 2 losing, so it is actually bad news for stock 2. Hence, even though the fundamentals of the two stocks are completely unrelated, the behavior of the index fund induces a negative correlation for the stock returns.

When stock 1 is in the index and the index fund holds most of its supply, the other investors' portfolios will largely consist of stock 2. The price of stock 2 at $t = 1$ includes a premium for the possibility that stock 2 is added to the index next period. If it is not added to the index at $t = 2$, this is bad news for the investors holding stock 2 and it reduces their realized returns. But given the negative cross-correlation between stocks 1 and 2, the investors can take a long position in stock 1 to hedge the index selection risk of their long position in stock 2. This creates a hedging demand for the index stock, driving its price up and expected return down, resulting in a negative expected return when the index fund is large.

**Index Fund Size and Risk Aversion**

Now we vary the size of the index fund ($h$) and the coefficient of absolute risk aversion ($b$) of the representative investor. We choose the following parameter values for the analysis:

\[
\begin{align*}
\mu &= 1 \\
\sigma &= 0.2 \\
\pi_{11} &= 1 \\
\pi_{21} &= 0.9 \\
h &\in [0, 1] \\
b &\in [0, 20]
\end{align*}
\]

With these parameter values, stock 1 is in the index at $t = 1$.

We vary risk aversion from zero (the risk-neutral case) to 20. The size of the index fund again varies from 0 to 1.

When the investors are risk-neutral, the stock prices are simply equal to the expected terminal payoffs. As the coefficient of absolute risk aversion increases, the price of stock 1 decreases. The smaller the index fund, the more the other investors will need to hold the stock
in equilibrium, and this drives down the price of stock 1.

The expected return for each stock starts at zero for the risk-neutral case regardless of the size of the index fund. For stock 1, which is in the index at \( t = 1 \), a larger index fund drives down the expected return. The expected return generally increases with risk aversion, except for a brief interval for a large index fund where the expected return actually decreases with aversion.

The price of stock 2 tends to increase slightly as the index fund grows, but in some middle regions it can even decrease and then increase again. The expected return on stock 2 also first increases and then decreases with \( h \).

While perhaps interesting to observe, these small nonmonotonicities remain small enough not to create concerns about the robustness of our results to the choice of the risk aversion coefficient.

### 3.4.2 Exogenous Index Rules

We now investigate the equilibrium when the index selection is random but independent of the dividend shocks and stock prices. At \( t = 2 \), we randomly choose which stock will be in the index that period, assigning probability \( q \) to stock 1 and probability \( 1 - q \) to stock 2. Note that when the index is selected exogenously, the index fund effectively becomes a regular large trader whose trading has a price impact but cannot be predicted by other investors.

The equilibrium prices at \( t = 2 \) are again given by:

\[
\begin{align*}
p_{12} &= \pi_{11} + \pi_{12} + \mu - (1 + \rho) b \sigma^2 + \rho h b \sigma^2 + 1_{\{1\}} (1 - \rho) h b \sigma^2 \\
p_{22} &= \pi_{21} + \pi_{22} + \mu - (1 + \rho) b \sigma^2 + \rho h b \sigma^2 + 1_{\{2\}} (1 - \rho) h b \sigma^2
\end{align*}
\] (3.33)

Similarly, the prices at \( t = 1 \) are also given by the same equations (3.26), (3.27), and (3.28) as before, except that now we calculate the expectations under an exogenous index selection rule.

Assuming without loss of generality that stock 1 is in the index at \( t = 1 \), we get the following
equilibrium prices (see the Appendix):

\[
\begin{align*}
    p_{11} &= \pi_{11} + 2\mu - 2(1 + \rho) b\sigma^2 + 2\rho\hat{b}\sigma^2 + (1 - \rho) h\hat{b}\sigma^2 + \\
         &\quad + (1 - \rho) h\hat{b}\sigma^2 \frac{q \exp \left(-b^2\sigma^2 h (1 - h)\right)}{q \exp \left(-b^2\sigma^2 h (1 - h)\right) + (1 - q) \exp \left(-b^2\sigma^2 h\right)} \\
    p_{21} &= \pi_{21} + 2\mu - 2(1 + \rho) b\sigma^2 + 2\rho\hat{b}\sigma^2 + \\
         &\quad + (1 - \rho) h\hat{b}\sigma^2 \frac{(1 - q) \exp \left(-b^2\sigma^2 h\right)}{q \exp \left(-b^2\sigma^2 h (1 - h)\right) + (1 - q) \exp \left(-b^2\sigma^2 h\right)}
\end{align*}
\] (3.35) (3.36)

The price of each stock is equal to its expected terminal payoff minus twice the one-period risk-premium \((1 + \rho) b\sigma^2\) plus twice the one-period shared index premium \(\rho\hat{b}\sigma^2\) plus the index premium \((1 - \rho) h\hat{b}\sigma^2\) for the current index stock plus the index premium for the possibility that each stock will be in the index in the future (i.e. \((1 - \rho) h\hat{b}\sigma^2\) times the risk-neutral probability of that stock being in the index next period).

To facilitate comparisons with the previous graphs, we choose the index stock at \(t = 1\) based on market capitalization exactly as before (i.e. stock 1 is in the index for \(\pi_{21} < 1\), and stock 2 for \(\pi_{21} \geq 1\)), so only the index selection at \(t = 2\) will be different.

Looking at the graphs for stock 1 across the \(\pi_{21}\) axis, we notice that the price and return of stock 1 are unaffected by the initial dividend shock \(\pi_{21}\) except for the cutoff at \(\pi_{21} = 1\) which determines the index composition at \(t = 1\). This was to be expected since the future index selection is now completely exogenous and unrelated to the dividend shocks.

The effect of index uncertainty can be observed as we vary the parameter \(q\) from 0 to 1. When \(q = 0.5\), there is maximum uncertainty about the index, since either stock can be in the index at \(t = 2\) with equal probabilities. Here the cross-correlation of stock returns is close but not exactly equal to zero, because there is still a small negative correlation induced by the random index selection – when one stock is chosen for the index, the other one is not chosen. This also creates a hedging demand for the index stock and decreases its expected returns. But now the index selection is not something that reinforces news about fundamentals, and therefore the effect remains much smaller than when the index selection was based on market cap.

When \(q = 1\), the index selection is completely predictable: stock 1 will be in the index at \(t = 2\) for sure. This eliminates the cross-correlation across the two stocks. The elimination of
the index uncertainty also reduces the variance of both stocks compared with the $q = 0.5$ case.

3.4.3 Portfolio Performance

How well does the index fund do under these different index selection rules? Is there some optimum level and type of index uncertainty for the index fund?

We choose the certainty equivalent of an investment opportunity between $t = 1$ and $t = 2$ as the performance measure of the index fund. It is defined as

$$u(W_1 + W_{CE}) = E_1[u(W_1 + W)]$$  \hspace{1cm} (3.37)

where $W_1$ denotes wealth at $t = 1$, $W_{CE}$ is the certainty equivalent, and $W$ is the random payoff on a zero net investment portfolio held from $t = 1$ to $t = 2$. The risk-free rate is zero so it does not show up in the equation. The certainty equivalent is determined using the utility function of the non-index investor, i.e. CARA utility with a coefficient of absolute risk aversion $b = 1$.

Alternatively, we could have used the Sharpe ratio as a performance measure. In this discrete-time model there is virtually no difference between the Sharpe ratio and the certainty equivalent as performance yardsticks. However, the Sharpe ratio ignores higher moments of the return distribution and generally in dynamic models it ignores the effects arising from changing investment opportunities, so we prefer to use the certainty equivalent.

It turns out that the certainty equivalent of the index stock (and thus that of the index fund) is more or less the same for both extremes: the market-cap index rule and the exogenous random selection rule (where $q = 0.5$). In the market-cap rule there is relatively little uncertainty about the index when the fundamentals of the two stocks differ very much, and this reduces the probability that the index fund will have to sell its portfolio at a low price at $t = 2$. However, in equilibrium the other investors anticipate this, and the greater certainty about the index composition makes them more willing to bid up the price of the index stock at $t = 1$. As a result, the index fund has to pay a higher index premium at $t = 1$. This effect exactly offsets the benefits of not having to trade again at $t = 2$.

However, the certainty equivalents are not identical across the selection rules. When the index is selected exogenously, there is still some hedging demand for the index stock, but this
demand is very low and almost exactly uniform across the initial dividend shocks. When the index is selected endogenously, the hedging demand is much greater, since now bad dividend news are also associated with bad index selection news. This effect is especially visible when there is a high probability of an index change, namely when the fundamentals of the stocks are close to each other. Therefore, the hedging demand drives down the expected return for the market-cap index especially for the middle values of $\pi_{21}$, and this hurts the performance of the index fund. This makes the exogenous index selection slightly more attractive than the market-cap rule to the index fund.

What about the situation where there is absolutely no uncertainty about the index composition at $t = 2$? We can do this within the exogenous index selection setup by choosing $q = 1$ (or equivalently $q = 0$) and comparing that with the $q = 0.5$ case. The certainty equivalents for the index stock (stock 1 or stock 2, depending on the value of $\pi_{21}$) for the two values of $q$ look very much alike. They decrease to approximately -0.2 ("overinvestment" in the index) as the index fund grows, and they are not affected by which stock is in the index at $t = 1$. However, there is a small difference: the certainty equivalent is slightly higher, especially for a large index fund, when there is no uncertainty about the index. This again is driven by the hedging demand induced by the negative correlation due to the index selection uncertainty. The difference in certainty equivalents is much smaller though than when comparing either of these cases with the market-cap rule, since the market-cap rule induced a much higher negative correlation and consequently a much higher hedging demand.

It should be noted that the difference in certainty equivalents is indeed due to the hedging demand effect and not due to a different trading behavior by the index fund. Even if stock 1 will be in the index for sure at $t = 2$, it does not matter for the performance of the index fund whether it buys stock 1 or stock 2 at $t = 1$. The price of each stock will adjust perfectly at $t = 1$ to accommodate the rational expectations about the index composition at $t = 2$.

What would be the best index selection rule from the point of view of the index fund? It turns out the best one would be a "reverse market-cap" rule where the stock with the smaller market cap would be selected for the index.\footnote{Reverse market-cap indices actually do exist, and they are tracked by index funds. Examples would include the Wilshire 4500 Index (total U.S. stock market minus S&P 500) and the Russell 2000 around its upper boundary (largest 3,000 firms minus largest 1,000).} Contrary to the usual market-cap rule, in this case
the index fund would buy the stock that experienced a sufficiently large negative fundamental shock. This positive demand shock would offset the negative fundamental shock and reduce the volatility of both stocks. The behavior of the index fund would in fact induce a positive correlation between the two stocks as good fundamental news for one stock would be good index news for the other stock. This in turn would create negative hedging demand for the index stock and increase its certainty equivalent in equilibrium.

Thinking about the index fund as a collection of several small index funds, we can isolate two externalities here. The small index funds would have a negative “price externality” on each other, since they would all buy the same stock at the same time and thus each index fund would move prices not only against itself but also against the other index funds. When index selection depends on market cap, the index funds would also have a “correlation externality” on each other – their collective behavior would shift the index premium from one stock to another in a way that either reinforces fundamental shocks (regular market-cap) or opposes fundamental shocks (reverse market-cap), inducing either negative or positive correlation between the two stocks and thereby either decreasing or increasing the certainty equivalent of the index stock. While the price externality of index funds on each other will always be negative, the correlation externality can be either positive or negative. In the reverse market-cap case, the positive correlation externality offsets some of the negative price externality that the index funds have on each other.

In summary, the rational anticipation of an index change cancels out any first-order effects on portfolio performance across the different index rules. The frequency of a deterministic index change does not matter at all. The probability of a random index change does matter a little. What matters by far the most is the choice between exogenous and endogenous index rules, especially if the probability of an index change is high. This difference becomes particularly visible as the index fund gets large, although for parameter values realistic for today’s U.S. stock market it might still remain economically small.
3.4.4 Robustness of Results

CARA Utility

While constant absolute risk aversion makes the model analytically convenient, it is of course not the most plausible utility function for a typical investor. Would our results change if we had some other utility function which exhibits decreasing absolute risk aversion?

One way to perform sensitivity analysis would be to solve the model for a convenient special case of DARA, such as constant relative risk aversion (CRRA). However, this is not possible in the current version of the model, since the terminal payoffs can theoretically be negative (this follows from assuming normal distribution which was very convenient with CARA utility) while CRRA utility is only defined for positive levels of wealth. Thus we need to use DARA utility functions which are defined for the whole real line \((-\infty, \infty)\). Deriving convenient analytical expressions in this case is complicated, but in fact we can get the direction of the main result without such laborious calculations.

Let us consider a CARA utility function which is perturbed slightly so that it now exhibits decreasing absolute risk aversion. This could be done e.g. by adding a function \(-\varepsilon F(W_{i3})\) to the investor's coefficient of absolute risk aversion \(b\), where \(\varepsilon > 0\) is a small constant and \(F(\cdot)\) is any cumulative distribution function (i.e. monotonically increasing from 0 to 1 on the interval \((-\infty, \infty)\)). We then get the following equation which would allow us to recover the corresponding utility function \(u(W)\):

\[
-\frac{u''(W)}{u'(W)} = b - \varepsilon F(W).
\]  

\[
u(W) = \int \exp \left( \int (\varepsilon F(W) - b) dW \right) dW
\]  

This is our starting point, although we proceed only with an intuitive explanation of the effects.

At \(t = 2\), when \(\rho = 0\) the stock prices are approximately given by (3.18) and (3.19), except that \(b\) is now greater or smaller, depending on the wealth of the mean-variance investor at that time. If the investor received a bad shock to the fundamentals at \(t = 1\), particularly to the non-index stock in which he has a larger position, he will be more risk-averse at \(t = 2\) which will increase \(b\) and thus decrease stock prices at \(t = 2\). Conversely, if the fundamental shocks
are good at $t = 1$, the investor will be less risk-averse at $t = 2$ and the prices will be higher. The effect of changing risk aversion therefore amplifies the volatility of the stock prices, and this reduces the prices of both stocks already at $t = 1$.

A negative fundamental shock to the non-index stock is particularly bad for the investor for two reasons: he has a larger position in that stock relative to the index stock, and under a market-cap rule the negative fundamental shock also makes it less likely that the non-index stock will be added to the index at $t = 2$. This increases the investor’s hedging demand for the index stock and pushes up the price of the index stock. The index fund will therefore suffer more from the market-cap rule than in the case with CARA utility.

**Identical Stocks**

In our model we have two stocks which have the same supply and which are identical in terms of their statistical properties. Perhaps our near-irrelevance result is not that surprising, given that we are just picking different ways to choose among fundamentally identical stocks.

The assumption of the two stocks being identical has been made so that we could isolate the effect of the index selection rule from any other competing effects. For example, if one of the stocks has a much larger supply, then the index fund will have a larger price impact on the small stock than on the large stock, since the fraction $h$ of the shares bought by the fund would no longer be essentially the same for the two stocks. This effect would favor a regular market-cap rule. But adding these other considerations would make it nearly impossible to identify the effect of the index selection rule alone, which is an interesting question in itself. Besides, regardless of the index selection rule, we could always let the index consist of a sufficiently large fraction of the stock market value (e.g. randomly select 70% of stock market cap, or select 70% of the market cap starting from the smallest firms and moving up). So there is no such inherent problem here as even a reverse market-cap rule can be implemented in a sensible way.

**Linear Price Impact**

Our CARA-normal structure produces an index premium that increases linearly with the size of the index fund. Perhaps this simple linear price impact of the index fund is what generates the near-irrelevance result.
In fact, it turns out linearity is not behind the result. At each point in time, the index stock will carry 100% of the index premium of that period. It will also carry part of the expected future index premium, i.e. the full index premium multiplied by its risk-neutral probability. What matters are the size of the index premium and how long the index stock is expected to have that premium – it does not matter whether the premium arose linearly or nonlinearly. Linearity would matter if the index premium could be split between the two stocks, but instead exactly one of the two stocks will be in the index at any one time.

### 3.5 Extension to Continuous Time

#### 3.5.1 Motivation

In this section we build a continuous-time extension of our model. This serves several purposes. The continuous-time methodology increases the analytical convenience of our results, in particular providing continuous analytical expressions for the instantaneous drift and the volatility of each stock with respect to each fundamental shock. A continuous-time setting makes it relatively easy to investigate the price dynamics over time, while going back one more period in the discrete-time setting (to the prices at \( t = 0 \)) would have been very complicated. Finally, this setting acts as a further robustness check for the results we obtained in our simple and intuitive discrete-time setting.

#### 3.5.2 Model Setup

The model is designed to be as close as possible to its discrete-time counterpart discussed before. We will analyze the case where index selection is based on market capitalization.

We have 2 stocks which each pay a dividend at the terminal date \( T \). The terminal dividend for stock \( i \) is given by:

\[
D_{iT} = \mu T + \sigma B_{iT} \quad i \in \{1, 2\}
\]  

where \( B_{it} \) is a standard Brownian motion at time \( t \), while \( \mu \) and \( \sigma \) are constants as before. We assume the dividend processes for the two stocks are uncorrelated. This could be generalized to allow nonzero correlation, but since for illustrative purposes we will perform the analysis of the equilibrium without correlation in fundamentals, we will also solve the equilibrium without
correlation to keep the equations as simple and intuitive as possible. The supply of each stock is 1.

There is a risk-free asset that exists in infinitely elastic supply. Its rate of return is fixed exogenously at \( r = 0 \).

At each point in time, the stock with the higher stock price will be chosen as the index. Since the fundamental dividend processes are the same for each stock, this is equivalent to saying that the stock with the higher currently observable dividend \( D_{it} \) will be the index. It turns out that both stock prices are continuous Ito processes – there are no discrete jumps in prices, even though the index selection is always a discrete choice. Any discontinuity in prices associated with index addition is ruled out by the absence of arbitrage. The index is updated continuously. The index fund will always hold a fraction \( h \) of the index stock.

The representative non-index investor maximizes CARA utility over terminal wealth.

### 3.5.3 Equilibrium

In equilibrium, we require that at each time \( t \in [0, T] \) the index fund holds a fraction \( h \) of the index stock (the one with the higher stock price), the representative CARA investor holds an optimum portfolio that maximizes his expected utility from terminal wealth, and the stock market clears. The market is complete, as there are two tradeable stocks and two sources of uncertainty.

The investor's problem is:

\[
\max_{\{a_{1t}, a_{2t}\}} \quad E[- \exp(-bW_{tT})] \\
\text{subject to} \quad W_{t0} \geq E[\pi_T W_{tT}] \tag{3.41}
\]

where \( \pi_t \) is the (unique) state-price density process at \( t \) normalized so that \( \pi_0 = 1 \). Denoting

---

\(^4\)Assume there is a discrete price jump \( \delta \) at the index boundary. Then at the boundary we can buy the non-index stock and short the index stock. If the non-index stock experiences a worse fundamental shock than the index shock and moves a distance \( \varepsilon \) from the boundary, we liquidate our position with a loss of \( \varepsilon \) per share. If the non-index stock experiences a better fundamental shock, it is added to the index, and we can liquidate our position with a gain of \( \delta + \varepsilon \) per share. As \( \varepsilon \to 0 \), this converges to an arbitrage opportunity (we have a positive probability of a positive future cashflow while never incurring any negative cashflow).
the Lagrange multiplier by \( \lambda \), the first-order condition is

\[
b \exp (-bW_{iT}) - \lambda \pi_T = 0. \tag{3.43}
\]

This determines the terminal state-price density as a function of terminal wealth. In equilibrium, the investor’s terminal wealth is equal to:

\[
W_{iT} = \left(1 - 1_{\{D_{iT} \geq D_{T}\}} h\right) D_{iT} + \left(1 - 1_{\{D_{iT} > D_{t}\}} h\right) D_{iT} + \Lambda_T + [W_{i0} - (1 - 1_{\{s_{10} > s_{20}\}} h) S_{10} - (1 - 1_{\{s_{20} > s_{10}\}} h) S_{20}]. \tag{3.44}
\]

In other words, the investor holds \( 1 - h \) of the index stock and \( 1 \) unit of the non-index stock. The remaining wealth of the investor is invested in the risk-free asset, earning a zero rate of return.

The somewhat surprising term in the expression is \( \Lambda_T \), which is the local time of the process \( S_{1t} - S_{2t} \) at 0. It represents the amount of cash generated by the strategy, i.e. the deviation of the investor’s strategy from a self-financing strategy. Initially it may seem that the investor’s trading strategy is self-financing: the stock price processes are continuous even around the index cutoff, and the investor simply shifts instantaneously from one stock to another as the dividends and thus also the stock prices cross. However, the Brownian motion processes in fact cross an infinite number of times, and since the investor always ends up selling the more expensive stock, thus gaining an infinitesimal dollar amount infinitely often, this trading around the index cutoff actually generates a positive cash flow. The same phenomenon has been discussed in the context of option pricing by Carr and Jarrow (1990).

The presence of local time considerably complicates the analytical solution, since local time depends on the price processes \( S_{1t} \) and \( S_{2t} \) but we cannot solve for the price processes until we have determined the state-price density, whereas the state-price density in turn depends on local time through the investor’s terminal wealth \( W_{iT} \). Hence, we conjecture that the effect of local time on pricing is “small,” and we proceed to compute a trial solution where local time does not matter for pricing. It turns out that our solutions are remarkably close to the discrete-time results.

Writing the wealth process as the sum of a constant \( c \) and the two dividends, we obtain the
state-price density at \( t \) as a conditional expectation:

\[
\pi_t = E_t[\pi_T] = \frac{b \exp\left(-bc\right)}{\lambda} E_t \left[ \exp \left(-b \left((1 - 1_{D_1 \geq D_2}) h\right) D_1 T + (1 - 1_{D_2 > D_1}) h\right) D_2 T \right].
\]

Knowing the state-price density, we can derive the stock price from the terminal dividend:

\[
S_{it} = \frac{E_t[\pi_T D_{it}]}{\pi_t}.
\] (3.47)

The only computational challenge here is calculating the expectations in the numerator and the denominator. The difficulty arises from the indicator function which breaks down the normal distribution structure and introduces a discreteness into the distribution. Nevertheless, after some tedious algebra we obtain:

\[
S_{it} = \mu T + \sigma B_{it} - b \sigma^2 \left(1 - \frac{h}{2} \right) (T - t) + \frac{b \sigma^2 h}{2} (T - t) \exp \left(b h \sigma B_{1t} \right) \left[1 - \Phi \left(A_{1t}\right)\right] - \exp \left(b h \sigma B_{2t} \right) \Phi \left(A_{2t}\right) + \exp \left(b h \sigma B_{2t} \right) \left[1 - \Phi \left(A_{2t}\right)\right] + \exp \left(b h \sigma B_{2t} \right) \Phi \left(A_{2t}\right)
\]

\[
+ \sigma \sqrt{\frac{T - t}{2}} \exp \left(b h \sigma B_{1t}\right) \phi \left(A_{1t}\right) - \exp \left(b h \sigma B_{2t}\right) \phi \left(A_{2t}\right) - \exp \left(b h \sigma B_{2t}\right) \Phi \left(A_{2t}\right),
\] (3.48)

where we have defined:

\[
A_{1t} = \frac{B_{2t} - B_{1t}}{\sqrt{2(T - t)}} - b h \sigma \sqrt{\frac{T - t}{2}},
\] (3.49)

\[
A_{2t} = \frac{B_{2t} - B_{1t}}{\sqrt{2(T - t)}} + b h \sigma \sqrt{\frac{T - t}{2}}.
\] (3.50)

\( \phi (\cdot) \) and \( \Phi (\cdot) \) denote the probability density function and the cumulative distribution function of the standard normal distribution. Due to symmetry, the stock price \( S_{2t} \) is given by the same expression, except that we need to switch \( B_{1t} \) and \( B_{2t} \).
We apply Ito’s lemma to derive the price dynamics for stock $i$ in the form:\textsuperscript{5}

$$dS_t = \mu dt + \sigma_i dB_t + \sigma_i dB_{2t}. \quad (3.51)$$

The partial derivatives that give us the drift and the diffusion terms turn out to be long expressions that seem to offer little intuition. Perhaps the best way to illustrate them is to show them in 3D graphs.

3.5.4 Analysis of Equilibrium

To keep our analysis comparable to the discrete-time setting, we pick the same parameter values:

$$t = 1$$

$$T' = 3$$

$$b = 1$$

$$\mu = 1$$

$$\sigma = 0.2$$

$$B_{1t} = 0$$

$$h \in [0,1]$$

$$B_{2t} \in [-2.5, 2.5]$$

Note that varying the second standard Brownian motion $B_{2t}$ from $-2.5$ to $2.5$ at time $t = 1$ corresponds to varying $\pi_{21}$ (the dividend shock to stock 2 at $t = 1$ in the discrete-time model) between 0.5 and 1.5, because $B_{2t}$ will be multiplied by its standard deviation $\sigma = 0.2$ and the nonzero mean of the dividend shocks is accounted for separately.

Now the prices and drift terms of both stocks appear much smoother than in the discrete-time case. This is not surprising, because in the continuous-time case under the market-cap rule, prices have to be continuous even around index addition. The overall shapes of prices,

\textsuperscript{5} Again we work with dollar returns $dS_t$ instead of rates of return $\frac{dS_t}{S_t}$. 

132
drift terms, volatility, and cross-correlation are the same as before. The biggest difference in these graphs is a higher negative cross-correlation (reaching -0.16 instead of -0.11). This arises from the infinitely fine time intervals as opposed to the lumpy time intervals and less frequent index updating of the discrete-time setup.

3.5.5 Portfolio Performance

The instantaneous certainty equivalent of stock 1 follows a pattern similar to the discrete-time case, displaying a sudden jump around index selection at $B_{2t} = 0$. This might be somewhat surprising since the stock price and its drift and volatility (and hence the Sharpe ratio) are smooth around index selection. The discontinuity in certainty equivalents arises from changing future investment opportunities, and it is needed to generate the stark portfolio choice of the non-index investor around the index cutoff.

The certainty equivalent of the index stock decreases from 0.020 in the absence of an index fund ($h = 0$) to a minimum of about -0.031 ($h = 1$). The lowest point is reached at $B_{2t} = 0$ when there is a maximum level of uncertainty about the future index composition. This additional index selection volatility hurts the index fund which is constrained to hold a certain supply of the index stock. Conversely, the non-index stock actually has a higher certainty equivalent for a higher $h$, because the non-index investors will be compensated in equilibrium for the additional volatility they must bear. Under a deterministic index selection rule, the certainty equivalent decreases from 0.2 to -0.2 as $h$ increases from 0 to 1, so the maximum difference from the market-cap rule is almost twice the difference in the discrete-time case. Hence, in the continuous-time setting the market-cap rule seems to hurt the index fund more than in the discrete-time setting.

3.6 Interpretations and Extensions

3.6.1 Interpretation of Results

We can think about our model as a very stylized representation of the S&P 500 index (the index stock) and the rest of the stock market (the non-index stock). These two portfolios would have a positive but imperfect correlation – e.g. the index portfolio could have a negative
loading on the SMB factor of Fama and French, while the non-index portfolio would have a positive loading on it. Both the index premium this period and the expectations of future index premia would be capitalized in the stock prices today. Based on the index rule, one of the two portfolios would be chosen to be the index each period. There would be some passive investors who simply buy the S&P 500 index, perhaps because that would be perceived as the lowest-cost way of gaining exposure to most of the stock market, and there would be some more sophisticated and active investors who can invest in both index stocks and non-index stocks.

The model suggests that the index funds will suffer when they follow a commonly tracked index, but the degree to which the funds are hurt does not depend that much on the randomness or predictability of the index selection rule. Hence, whether the S&P 500 is updated frequently or not should already be reflected in the index premium, and consequently the expected index turnover cost paid by the index funds would not be affected by the frequency of index changes. The performance of the Russell 1000 might potentially be a little worse than that of the S&P 500, because the Russell 1000 is mechanically based on market cap unlike the S&P 500, but the difference would probably be swamped by other relevant considerations in the real market. In fact, the main drag on expected returns would be the price externality of the other indexers, and since the Russell 1000 is much less commonly tracked than the S&P 500, it should have a smaller price premium and tracking it should in this sense be a better deal for the index funds. In addition to the current level of the index premium, also its near-term movement up and down caused by future index fund inflows and outflows is relevant to index investors, at least to the extent that those fund flows are not already anticipated and reflected in today’s market prices.

It should be noted that our model provides a benchmark result in the presence of frictionless markets and rational investors. Both of these assumptions are of course simplifications of reality. They constitute a good starting point, but before we apply the prescriptions of the model to the real stock market, we need a few caveats. How much would transaction costs influence the performance of index funds? And is some of the actual index premium a result of irrational mispricing which might also change the way this premium behaves across different index rules? Transaction costs should be a conceptually simple extension to our model. But how would we model an irrational index premium? What are the laws governing this irrational
phenomenon? It would require extensive empirical research into the index premium to establish its predictable deviations from rationality under different index rules.

3.6.2 Index Effect for a Large Number of Stocks

The NYSE, AMEX, and Nasdaq have about 8,000 listed stocks, yet our model illustrates the index effect for two stocks or two portfolios of stocks. How exactly would the index premium behave at the level of an individual stock when the total number of stocks is large? Using the two-stock model as a benchmark, this question is relatively easy to answer. The rational index premium arises when the non-index investors cannot achieve full diversification benefits without the index stocks.

Just like in the model, we assume there are two types of investors: index investors who buy only the index stocks, and non-index investors who are rational and unconstrained mean-variance optimizers. We also assume the non-index investors have full information about the means, variances, and correlation structure of returns across stocks. We consider a static one-period model where each stock return is generated by a $k$-factor model (i.e. $R_i = E_i + \sum_{j=1}^{k} \beta_{ij} f_j + \epsilon_i$). For simplicity, we let each stock have a market beta equal to 1, but the stocks differ in their other factor loadings and their idiosyncratic risk.

When there are no index investors, the mean-variance investors pick mean-variance efficient portfolios, and we obtain the standard CAPM result where everyone holds the market portfolio which is also mean-variance efficient.

Now let us say there are 1,000 stocks, and we randomly select 500 of them to be in the index. Let us also have some index investors who buy, say, 20% of each index stock. In equilibrium, the remaining portfolio of the 500 non-index stocks and 80% of the supply of the 500 index stocks will now have to be mean-variance efficient. To make it mean-variance efficient, the prices of the index stocks will have to increase slightly relative to the non-index stocks, so that the non-index stocks have slightly more attractive expected returns than the index stocks. The difference in the prices of index stocks between this case and the standard CAPM case is the index premium.

However, the size of the index premium in this case will be negligible. The 500 index stocks form a randomly selected well-diversified portfolio, and this portfolio will therefore have
an almost perfect correlation with the market factor. Similarly, the 500 non-index stocks constitute a randomly selected well-diversified portfolio which will also have an almost perfect correlation with the market factor. These two portfolios are almost perfect substitutes for each other, just like in our two-stock model as $\rho \to 1$. Hence, even a tiny change in the expected return on one of them will induce the mean-variance investors to tilt their holdings to favor one portfolio over the other. Consequently, even an extremely small price premium on index stocks will be enough to induce the non-index investors to hold market-clearing quantities of stocks. As the number of stocks increases, the index premium converges to zero. The demand curve for an individual stock becomes horizontal.

If instead the index stocks have on average a different exposure to a systematic risk factor than the non-index stocks, then the index portfolio and the non-index portfolio will no longer be perfectly correlated (as in the model when $0 \leq \rho < 1$). Thus the non-index investors are no longer indifferent between holding index stocks and non-index stocks – they prefer to hold both index stocks and non-index stocks for diversification reasons. In our rational asset-pricing setting, a clearly observable index premium can only be created by a difference in systematic risk between the index stocks and non-index stocks.

For the S&P 500 index where market capitalization plays an important role, such a systematic factor risk could be something like the SMB (small minus big) factor of Fama and French. The large-cap index stocks would have a negative loading on SMB while the non-index stocks would generally have a positive loading on it. Holding the market portfolio would eliminate all SMB exposure (in our setting, SMB would be uncorrelated with the market portfolio), so excluding some large-cap index stocks with negative SMB loadings will produce a portfolio with a positive SMB loading. To offset their positive SMB factor loading, the non-index investors would have some demand for the index stocks in particular, and this would create an index premium.

Fundamentally this index premium would actually not be an index premium per se but a premium on one of the systematic risk factors. By buying the index stocks which would generally have a negative loading on SMB, the index investors would indirectly exhibit a preference for a negative SMB loading, thus creating a price premium on stocks with a negative SMB loading. This price premium would in fact be identical for index stocks and non-index stocks
with a similar SMB loading – those few index stocks with a positive SMB loading would be trading at a discount, while the non-index stocks with a negative SMB loading would carry the same price premium as similar index stocks.

Would the price of an individual stock jump when it is added to the index? The answer is no. Index membership does not affect the future cash flows nor the systematic risk of a firm, and the demand curve for an individual stock would still be horizontal. The price of a stock could jump up only if index addition somehow changed the factor loadings of the stock.

If we allow for incomplete information or institutional rigidities, other possibilities emerge. Investors might view index membership as a proxy for factor risk – e.g. they might trade in the S&P 500 stocks to adjust their perceived exposure to large-cap stocks. Although investors understand that factor risk does not change on the announcement day of index addition, they still consider the S&P 500 index a good first approximation to a portfolio that is exposed to large-cap factor risk. This kind of thinking in terms of categories could contribute to an index premium. In fact, Barberis, Shleifer, and Wurgler (2001) present some evidence suggesting that index addition could actually change the factor betas of a stock and that this would be driven by shifts in investors' demand for certain categories of assets. However, this explanation cannot create the index premium because it requires that noise trader flows can affect the prices of stocks directly, so it can only amplify the effect when something else created it in the first place.

To mitigate the index premium, one might be tempted to suggest picking the index more or less randomly (perhaps excluding some micro-cap or closely held stocks). Then the index would have no non-market factor exposure, and there would be no rational justification for an index premium. However, stock market indices are, by their very nature, collections of similar stocks such as large-caps, tech stocks, industry portfolios, etc., since the indices have primarily been designed to measure the performance of various segments of the stock market. Perhaps the only exception is the Wilshire 5000 index which covers basically all publicly traded U.S. firms. Stocks in an index do therefore share some characteristics and also more likely share some non-market factor risk than a randomly selected portfolio of stocks, giving some reason for the existence of a rational index premium.
3.7 Conclusions

When index funds are significant players in the stock market, they clearly cannot just free-ride on efficient market prices. The index funds themselves will move prices, which has also been well documented in the empirical literature. For an investor making a decision between a passive index fund and an actively managed fund, the transaction costs are not the only relevant costs to compare – the index fund will also include a systematic but somewhat hidden cost due to tracking a well-known index in which all the member stocks are trading at a premium.

In this chapter we have investigated the behavior of the index premium and the performance of index funds under different index selection rules. A priori it seems like the updating frequency or probability would affect the index turnover cost borne by the index funds, since it affects the frequency of paying the index premium as well as the size of the index premium. The index premium might also be affected by the type of the index selection rule, i.e. whether it is based on an endogenous random variable such as market capitalization or an exogenous random or deterministic event.

It turns out that the rational anticipation of the future index composition reflected in stock prices today eliminates any first-order differences in index fund performance across index selection rules. As the index fund becomes large, the other investors become less diversified, and this induces hedging motives which do create some interesting price effects. The most adversely affected indices would be the ones based on a market-cap rule. Yet these effects could still be economically small for reasonable parameter values for the U.S. stock market today.

These results should be interpreted as a benchmark case. In the presence of rational investors and frictionless markets, index funds should be close to indifferent to the index rule they are following, unless the index funds constitute a large part of the market. If we add in a specific form of transaction costs, the index funds might develop a strong preference for one rule over another, depending on the type of the transaction cost. Some empirically observed price effects associated with index addition and deletion also suggest that the market’s anticipation of future index changes might not have been perfectly rational, and this might also have implications for the performance of the index funds. Now that we have the benchmark result of this chapter, we can begin to examine the impact that these other considerations would have
on the optimal index selection rule and, consequently, on the comparison between active and passive money management.
3.8 References

Barberis, N. and A. Shleifer, 2000, "Style Investing," working paper, University of Chicago and Harvard University.


Goetzmann, W.N. and M. Massa, 1999, "Index Funds and Stock Market Growth," working paper, Yale School of Management and INSEAD.


3.9 Appendix: Formulas in Discrete Time

3.9.1 Equilibrium Prices at $t = 2$

The investor's objective function at $t = 2$ is:

$$\max_{a_{112}, a_{22}} E_2 \left[ -\exp \left(-bW_{13}\right) \right]$$

subject to:

$$W_{13} = W_{12} + a_{112} (p_{13} - p_{12}) + a_{22} (p_{23} - p_{22})$$

Since the terminal payoffs and thus the terminal prices $p_{13}$ and $p_{23}$ are normally distributed, the problem reduces to:

$$\max_{a_{112}, a_{22}} \mathbb{E}_2 [W_{13}] - \frac{1}{2} b \text{Var}_2 [W_{13}]$$

$$\Leftrightarrow \max_{a_{112}, a_{22}} a_{112} (E_2 [p_{13}] - p_{12}) + a_{22} (E_2 [p_{23}] - p_{22}) - \frac{1}{2} b \sigma_a^2 (a_{112}^2 + a_{22}^2 + 2a_{112}a_{22} \rho)$$

The first-order conditions with respect to stock holdings $a_{112}$ and $a_{22}$ are:

$$E_2 [p_{13}] - p_{12} - b \sigma_a^2 a_{112} - b \sigma_a^2 p_{112} = 0$$

$$E_2 [p_{23}] - p_{22} - b \sigma_a^2 a_{22} - b \sigma_a^2 p_{112} = 0$$

Solving for the stock holdings, we get:

$$a_{112} = \frac{1}{b \sigma_a^2 (1 - \rho^2)} \left[ E_2 [p_{13}] - p_{12} - \rho (E_2 [p_{23}] - p_{22}) \right]$$

$$a_{22} = \frac{1}{b \sigma_a^2 (1 - \rho^2)} \left[ E_2 [p_{23}] - p_{22} - \rho (E_2 [p_{23}] - p_{22}) \right]$$

The market-clearing conditions are:

$$a_{112} + a_{f12} = 1$$

$$a_{22} + a_{f22} = 1$$

The demand of the index fund is simply:

$$a_{f12} = \frac{W_{f2}}{p_{12}}$$

$$a_{f22} = \frac{W_{f2}}{p_{22}}$$

In other words, if stock 1 is in the index (represented by the event \{1\}), the index fund will spend all of its wealth $W_{f2}$ on stock 1, and if the stock 2 is in the index (event \{2\}), all of $W_{f2}$ is invested in stock 2. Using the market-clearing
conditions, we immediately get the equilibrium holdings of the investor:

\[ a_{12} = 1 - \frac{W_{12}}{p_{12}} \]
\[ a_{22} = 1 - \frac{W_{22}}{p_{22}} \]

In equilibrium these stock holdings have to be optimal for the investor, so we plug these expressions back into the first-order conditions. This gives us the equilibrium stock prices:

\[ p_{12} = \frac{1}{2} \left[ E_{3} [p_{13}] - (1 + \rho) \sigma^2 + \sqrt{(E_{2} [p_{13}] - (1 + \rho) \sigma^2)^2 + 4\sigma^2 W_{2} \left( \rho + 1 \right) (1 - \rho)} \right] \]
\[ p_{22} = \frac{1}{2} \left[ E_{3} [p_{23}] - (1 + \rho) \sigma^2 + \sqrt{(E_{2} [p_{23}] - (1 + \rho) \sigma^2)^2 + 4\sigma^2 W_{2} \left( \rho + 1 \right) (1 - \rho)} \right] \]

To simplify these expressions, we set the demand of the index fund equal to \( h \) units of the index stock, regardless of which stock is in the index. Then we get much more intuitive and convenient expressions for stock prices, and the equilibrium will be given by:

\[ a_{f12} = \frac{1}{2} h \]
\[ a_{f22} = \frac{1}{2} h \]
\[ a_{12} = 1 - \frac{1}{2} h \]
\[ a_{22} = 1 - \frac{1}{2} h \]
\[ p_{12} = E_{3} [p_{13}] - (1 + \rho) \sigma^2 + \rho h^2 + 1 \left( \frac{1}{2} \right) (1 - \rho) h \sigma^2 \]
\[ = \pi_{12} + \pi_{12} + \mu - (1 + \rho) \sigma^2 + \rho h^2 + 1 \left( \frac{1}{2} \right) (1 - \rho) h \sigma^2 \]
\[ p_{22} = E_{3} [p_{23}] - (1 + \rho) \sigma^2 + \rho h^2 + 1 \left( \frac{1}{2} \right) (1 - \rho) h \sigma^2 \]
\[ = \pi_{22} + \pi_{22} + \mu - (1 + \rho) \sigma^2 + \rho h^2 + 1 \left( \frac{1}{2} \right) (1 - \rho) h \sigma^2 \]

### 3.9.2 Equilibrium Prices at \( t = 1 \)

At \( t = 1 \) the objective function of the investor is:

\[
\max_{a_{11}, a_{12}, a_{22}} E_{1} \left[ -\exp(-bW_{13}) \right]
\]

subject to :

\[ W_{13} = W_{12} + a_{12} (p_{13} - p_{12}) + a_{12} (p_{23} - p_{22}) \]
\[ W_{12} = W_{13} + a_{11} (p_{12} - p_{11}) + a_{21} (p_{22} - p_{21}) \]
By iterated expectations, this can be expressed as:

\[
\max_{a_{t11}, a_{t12}, a_{t21}, a_{t22}} E_1 \left[ E_2 \left[ -\exp(-bW_{t3}) \right] \right]
\]

subject to:

\[
W_{t3} = W_{t2} + a_{t12} (p_{13} - p_{12}) + a_{t22} (p_{23} - p_{22})
\]

\[
W_{t2} = W_{t1} + a_{t11} (p_{12} - p_{11}) + a_{t21} (p_{22} - p_{21})
\]

Given the information available at \( t = 2 \), \( W_{t3} \) will be normally distributed because the only random components in it are the normally distributed dividends \( \pi_{t3} \) at \( t = 3 \). Therefore, we can write the objective function as:

\[
\max_{a_{t11}, a_{t12}, a_{t21}, a_{t22}} E_1 \left[ -\exp \left( -bE_2 \left[ W_{t3} \right] + \frac{1}{2} b^2 \text{Var}_2 \left[ W_{t3} \right] \right) \right]
\]

Anticipating the equilibrium outcome at \( t = 2 \), the conditional expectation and variance of \( W_{t3} \) will be:

\[
E_2 \left[ W_{t3} \right] = W_{t3} + \sigma^2 \left[ 1 + 2\rho (1 - h) + (1 - h)^2 \right]
\]

\[
\text{Var}_2 \left[ W_{t3} \right] = \sigma^2 \left[ 1 + 2\rho (1 - h) + (1 - h)^2 \right]
\]

The conditional variance of \( W_{t3} \) is constant, while the conditional expectation is a constant plus \( W_{t2} \). This simplifies the objective function to a utility maximization using wealth one period ahead (at \( t = 2 \)) rather than two periods ahead (at \( t = 3 \)):

\[
\max_{a_{t11}, a_{t21}} \exp \left[ -\frac{1}{2} b^2 \sigma^2 \left( 1 + 2\rho (1 - h) + (1 - h)^2 \right) \right] E_1 \left[ -\exp(-bW_{t2}) \right]
\]

\[
\equiv \max_{a_{t11}, a_{t21}} E_1 \left[ -\exp(-bW_{t2}) \right]
\]

However, given the information available at \( t = 1 \), \( W_{t2} \) is not normally distributed. Although the dividend shocks which determine \( p_{12} \) and \( p_{22} \) are normally distributed, these shocks also determine the composition of the index, and this introduces a discreteness into the distribution of prices at \( t = 2 \) given the information we have at \( t = 1 \). Hence, we can no longer apply the convenient CARA normal methodology where we would simplify the objective function to an expectation minus a variance – instead we need to compute it in a more tedious way.

To obtain the first-order conditions with respect to the stock holdings \( a_{t11} \) and \( a_{t21} \), we need to plug in the expressions for \( W_{t2} \), \( p_{12} \) and \( p_{22} \). After performing the substitutions and the differentiation, we get two equations for the two stock prices:

\[
0 = (\pi_{t1} + \mu - (1 + \rho) b \sigma^2 + \rho b h \sigma^2 - p_{t1}) E_1 \left[ \exp \left( -ba_{t11} \left( \pi_{t2} + 1_{t1} (1 - \rho) h \sigma^2 \right) - ba_{t21} \left( \pi_{t2} + 1_{t2} (1 - \rho) h \sigma^2 \right) \right) \right] + E_1 \left[ (\pi_{t2} + 1_{t1} (1 - \rho) h \sigma^2) \exp \left( -ba_{t11} \left( \pi_{t2} + 1_{t1} (1 - \rho) h \sigma^2 \right) - ba_{t21} \left( \pi_{t2} + 1_{t2} (1 - \rho) h \sigma^2 \right) \right) \right]
\]

\[
0 = (\pi_{t2} + \mu - (1 + \rho) b \sigma^2 + \rho b h \sigma^2 - p_{t2}) E_1 \left[ \exp \left( -ba_{t11} \left( \pi_{t2} + 1_{t1} (1 - \rho) h \sigma^2 \right) - ba_{t21} \left( \pi_{t2} + 1_{t2} (1 - \rho) h \sigma^2 \right) \right) \right] + E_1 \left[ (\pi_{t2} + 1_{t2} (1 - \rho) h \sigma^2) \exp \left( -ba_{t11} \left( \pi_{t2} + 1_{t1} (1 - \rho) h \sigma^2 \right) - ba_{t21} \left( \pi_{t2} + 1_{t2} (1 - \rho) h \sigma^2 \right) \right) \right]
\]
These equations can be solved for the prices in closed form. This yields:

\[
\begin{align*}
p_{11} &= \pi_{11} + \mu - (1 + \rho) b \sigma^2 + \rho b \sigma^2 + \\
&\quad + \frac{E_1 \left[ (\pi_{12} + 1_{(i)}) (1 - \rho) h b \sigma^2 \right] \exp \left( -ba_{111} \left( \pi_{12} + 1_{(i)} \right) (1 - \rho) h b \sigma^2 \right) - ba_{21} \left( \pi_{22} + 1_{(j)} \right) (1 - \rho) h b \sigma^2) \right]}{E_1 \left[ \exp \left( -ba_{111} \left( \pi_{12} + 1_{(i)} \right) (1 - \rho) h b \sigma^2 \right) - ba_{21} \left( \pi_{22} + 1_{(j)} \right) (1 - \rho) h b \sigma^2) \right]} \right] \\
p_{21} &= \pi_{21} + \mu - (1 + \rho) b \sigma^2 + \rho b \sigma^2 + \\
&\quad + \frac{E_1 \left[ (\pi_{22} + 1_{(j)}) (1 - \rho) h b \sigma^2 \right] \exp \left( -ba_{111} \left( \pi_{12} + 1_{(i)} \right) (1 - \rho) h b \sigma^2 \right) - ba_{21} \left( \pi_{22} + 1_{(j)} \right) (1 - \rho) h b \sigma^2) \right]}{E_1 \left[ \exp \left( -ba_{111} \left( \pi_{12} + 1_{(i)} \right) (1 - \rho) h b \sigma^2 \right) - ba_{21} \left( \pi_{22} + 1_{(j)} \right) (1 - \rho) h b \sigma^2) \right]} \right]
\end{align*}
\]

This allows us to define the pricing kernel (i.e. the state price density, or the stochastic discount factor) of the economy at \( t = 1 \) for payoffs at \( t = 2 \):

\[
m_2 = \frac{\exp \left( -ba_{111} \left( \pi_{12} + 1_{(i)} \right) (1 - \rho) h b \sigma^2 \right) - ba_{21} \left( \pi_{22} + 1_{(j)} \right) (1 - \rho) h b \sigma^2) \right]}{E_1 \left[ \exp \left( -ba_{111} \left( \pi_{12} + 1_{(i)} \right) (1 - \rho) h b \sigma^2 \right) - ba_{21} \left( \pi_{22} + 1_{(j)} \right) (1 - \rho) h b \sigma^2) \right]} \right]
\]

We then get the following general expressions for the stock prices:

\[
\begin{align*}
p_{11} &= \pi_{11} + \mu - (1 + \rho) b \sigma^2 + \rho b \sigma^2 + E_1 \left[ (\pi_{12} + 1_{(i)}) (1 - \rho) h b \sigma^2) \right] m_2 \\
p_{21} &= \pi_{21} + \mu - (1 + \rho) b \sigma^2 + \rho b \sigma^2 + E_1 \left[ (\pi_{22} + 1_{(j)}) (1 - \rho) h b \sigma^2) \right] m_2
\end{align*}
\]

This holds for all index selection rules. Next we compute the expectations for the market-cap rule and the exogenous randomization rule.

### 3.9.3 Equilibrium Prices under a Market-Cap Rule

Now \( \{1\} = \{p_{12} > p_{22}\} = \{\pi_{11} + \pi_{12} > \pi_{21} + \pi_{22}\} \), so the stock with the better fundamental shocks (and thus a higher stock price) will be chosen for the index at \( t = 2 \). We start by calculating the denominator of the state-price density (the subscript \( i \) which refers to the investor (as opposed to the index fund) is dropped from \( a_{ij} \) for simplicity):

\[
E_1 \left[ \exp \left( -ba_{111} \left( \pi_{12} + 1_{(i)} \right) (1 - \rho) h b \sigma^2 \right) - ba_{21} \left( \pi_{22} + 1_{(j)} \right) (1 - \rho) h b \sigma^2) \right] 
\]

The problem here is that index selection (the indicator functions \( 1_{(i)} \) and \( 1_{(j)} \)) depends on both dividend shocks \( \pi_{12} \) and \( \pi_{22} \). We perform a change of variables to make it dependent only on one of two independent fundamental shocks. Define:

\[
\begin{align*}
\bar{\pi}_{12} &= \pi_{12} - \pi_{22} \sim N \left( 0, 2(1 - \rho) \sigma^2 \right) \\
\bar{\pi}_{22} &= \pi_{12} + \pi_{22} \sim N \left( 2\mu, 2(1 + \rho) \sigma^2 \right)
\end{align*}
\]

Then we can substitute

\[
\begin{align*}
\pi_{12} &= \frac{\bar{\pi}_{12} + \bar{\pi}_{22}}{2} \\
\pi_{22} &= \frac{\bar{\pi}_{22} - \bar{\pi}_{12}}{2}
\end{align*}
\]

145
and further

\[ \tilde{\pi}_{12} = \sigma \sqrt{2(1 - \rho)} \mu_1 \]
\[ \tilde{\pi}_{22} = \sigma \sqrt{2(1 + \rho)} \mu_2 + 2 \mu \]

where \( \mu_1 \) and \( \mu_2 \) are distributed \( \mathcal{N}(0, 1) \). This allows us to compute the expectation of a product as a product of expectations:

\[
E_1 \left[ \exp \left( -\frac{b}{2} (a_{11} - a_{21}) \tilde{\pi}_{12} - a_{11} \frac{b^2}{2} \sigma^2 (1 - \rho) 1_{[1]} - a_{21} \frac{b^2}{2} \sigma^2 (1 - \rho) 1_{[2]} \right) \right] \times

\forall \left[ \exp \left( -\frac{b}{2} (a_{11} + a_{21}) \tilde{\pi}_{22} \right) \right]
\]

\[
= E_1 \left[ \exp \left( -\frac{b}{2} (a_{11} - a_{21}) \sigma \sqrt{2(1 - \rho)} \mu_1 - a_{21} \frac{b^2}{2} \sigma^2 (1 - \rho) \right) \right] \times

\forall \left[ \exp \left( -\frac{b}{2} (a_{11} + a_{21}) \mu_2 + 2 \mu \right) \right]
\]

\[
= \exp \left[ -b \mu (a_{11} + a_{21}) + \frac{1}{4} b^2 \sigma^2 (a_{11} + a_{21})^2 (1 + \rho) + \frac{1}{4} b^2 \sigma^2 (a_{11} - a_{21})^2 (1 - \rho) \right] \times

\forall \left[ \exp \left[ -a_{21} \frac{b^2}{2} \sigma^2 (1 - \rho) \right] [1 - \Phi(A)] + \exp \left[ -a_{21} \frac{b^2}{2} \sigma^2 (1 - \rho) \right] \Phi(A) \right]
\]

where we defined

\[
A = \frac{\pi_{21} - \pi_{11}}{\sigma \sqrt{2(1 - \rho)}} + b \sigma (a_{11} - a_{21}) \sqrt{\frac{1 - \rho}{2}}
\]

Since the equilibrium holdings \( a_{11} \) and \( a_{21} \) are always 1 and \( 1 - h \) (either way), the integral further reduces to:

\[
= \exp \left[ -b \mu (2 - h) + \frac{1}{4} b^2 \sigma^2 (2 - h)^2 (1 + \rho) + \frac{1}{4} b^2 \sigma^2 h^2 (1 - \rho) \right] \times

\forall \left[ \exp \left[ -a_{11} \frac{b^2}{2} \sigma^2 (1 - \rho) \right] [1 - \Phi(A)] + \exp \left[ -a_{21} \frac{b^2}{2} \sigma^2 (1 - \rho) \right] \Phi(A) \right]
\]

Next we calculate the first term in the numerator:

\[
E_1 \left[ \pi_{11} \exp \left( -b a_{11} (\pi_{12} + 1_{[1]} (1 - \rho) h \sigma^2) - b a_{21} (\pi_{22} + 1_{[2]} (1 - \rho) h \sigma^2) \right) \right]
\]

We do all the calculations for stock 1. Due to symmetry, we then immediately get the price of stock 2 as well. Again, we
perform the same substitutions for \( \pi_{12} \) and \( \pi_{22} \) and compute the integrals as before. This yields:

\[
\sigma \sqrt{\frac{1-\rho}{2}} \exp \left[ -b\mu (2-h) + \frac{1}{4} b^2 \sigma^2 (2-h)^2 (1+\rho) \right] \times \\
\times \left[ \exp \left[ -a_{11} b^2 \sigma^2 (1-\rho) \right] \left[ \phi(A) + b\sigma \left( a_{21} - a_{11} \right) \sqrt{\frac{1-\rho}{2}} (1-\Phi(A)) \right] + \\
+ \exp \left[ -a_{21} h b^2 \sigma^2 (1-\rho) \right] \left[ -\phi(A) + b\sigma \left( a_{21} - a_{11} \right) \sqrt{\frac{1-\rho}{2}} \Phi(A) \right] \right] + \\
+ \left( \mu - \frac{1}{2} b\sigma^2 (2-h) (1+\rho) \right) \exp \left[ -b\mu (2-h) + \frac{1}{4} b^2 \sigma^2 (2-h)^2 (1+\rho) + \frac{1}{4} b^2 \sigma^2 h^2 (1-\rho) \right] \times \\
\times \left[ \exp \left[ -a_{11} h b^2 \sigma^2 (1-\rho) \right] \left[ 1-\Phi(A) \right] + \exp \left[ -a_{21} h b^2 \sigma^2 (1-\rho) \right] \Phi(A) \right]
\]

The second term in the numerator gives us:

\[
E_1 \left[ 1_{1(1)} (1-\rho) h b \sigma^2 \exp \left( -b a_{11} \left( \pi_{12} + 1_{1(1)} (1-\rho) h b \sigma^2 \right) - b a_{21} \left( \pi_{22} + 1_{2(3)} (1-\rho) h b \sigma^2 \right) \right) \right] \\
= \left( 1-\rho \right) h b \sigma^2 \exp \left[ -b\mu (2-h) + \frac{1}{4} b^2 \sigma^2 (2-h)^2 (1+\rho) \right] \times \\
\times \left[ \frac{1}{4} h b^2 \sigma^2 h^2 (1-\rho) - a_{11} h b^2 \sigma^2 (1-\rho) \right] \Phi(A) + \\
\exp \left[ -b\mu (2-h) + \frac{1}{4} b^2 \sigma^2 (2-h)^2 (1+\rho) \right].
\]

Canceling out the term

\[
\exp \left[ -b\mu (2-h) + \frac{1}{4} b^2 \sigma^2 (2-h)^2 (1+\rho) \right]
\]

from the numerator and denominator, we finally get the price of stock 1:

\[
p_{11} = \pi_{11} + 2\mu - 2(1+\rho) b a_{11} + \frac{3}{2} b h b \sigma^2 + \frac{1}{2} h b \sigma^2 + \\
+ \frac{1}{2} b a_{21} - a_{11} (1-\rho) \exp \left[ \frac{1}{4} b^2 \sigma^2 h^2 (1-\rho) \right] + \\
+ (1-\rho) h b \sigma^2 \frac{\exp \left[ -a_{11} h b^2 \sigma^2 (1-\rho) \right] \left[ 1-\Phi(A) \right]}{\exp \left[ -a_{11} h b^2 \sigma^2 (1-\rho) \right] \left[ 1-\Phi(A) \right] + \exp \left[ -a_{21} h b^2 \sigma^2 (1-\rho) \right] \Phi(A)} \\
+ \sigma \sqrt{\frac{1-\rho}{2}} \exp \left[ \frac{1}{4} h b^2 \sigma^2 h^2 (1-\rho) \right] \frac{\left[ \exp \left[ -a_{11} h b^2 \sigma^2 (1-\rho) \right] - \exp \left[ -a_{21} h b^2 \sigma^2 (1-\rho) \right] \phi(A) \right]}{\exp \left[ -a_{11} h b^2 \sigma^2 (1-\rho) \right] \left[ 1-\Phi(A) \right] + \exp \left[ -a_{21} h b^2 \sigma^2 (1-\rho) \right] \Phi(A)}
\]

Depending on which stock is in the index at \( t = 1 \), we plug in 1 and 1 - \( h \) for \( a_{11} \) and \( a_{21} \) to obtain the stock price.

### 3.9.4 Equilibrium Prices under an Exogenous Index Rule

Now the index is selected exogenously, i.e. independently of the dividend shocks. The probability of stock 1 being in the index at \( t = 2 \) (event \( \{1\} \)) is \( q \). Now the index selection can be separated into its own expectation, which simplifies the
calculations considerably. Proceeding as before, we get:

\[ E_1 \left[ \exp \left( -\beta a_{11} \left( \pi_{12} + 1_{[1]} (1 - \rho) \beta b \sigma^2 \right) - \beta a_{21} \left( \pi_{22} + 1_{[2]} (1 - \rho) \beta b \sigma^2 \right) \right) \right] \\
= E_1 \left[ \exp \left( -\frac{b}{2} (a_{11} - a_{21}) \pi_{12} \right) \right] E_2 \left[ \exp \left( -\frac{b}{2} (a_{11} - a_{21}) \pi_{22} \right) \right] \times \\
\times E_1 \left[ \exp \left( -\frac{b}{2} (a_{11} + a_{21}) \pi_{22} \right) \right] \times \\
\times \exp \left[ -b_0 (a_{11} + a_{21}) + \frac{1}{4} b^2 \sigma^2 (a_{11} + a_{21})^2 (1 + \rho) + \frac{1}{4} b^2 \sigma^2 (a_{11} - a_{21})^2 (1 - \rho) \right] \times \\
\times \left[ \exp \left[ -a_{11} \beta b \sigma^2 (1 - \rho) \right] q + \exp \left[ -a_{21} \beta b \sigma^2 (1 - \rho) \right] (1 - q) \right].

We compute the numerator the same way. This gives us the stock prices:

\[ p_{11} = \pi_{11} + 2\mu - (1 + \rho) b \sigma^2 + \rho b \sigma^2 - b (a_{11} + \rho a_{21}) \sigma^2 + (1 - \rho) \beta b \sigma^2 + \frac{q \exp \left( -b^2 \sigma^2 h a_{11} \right)}{q \exp \left( -b^2 \sigma^2 h a_{11} \right) + (1 - q) \exp \left( -b^2 \sigma^2 h a_{21} \right)} \]

\[ p_{21} = \pi_{21} + 2\mu - (1 + \rho) b \sigma^2 + \rho b \sigma^2 - b (a_{21} + \rho a_{11}) \sigma^2 + (1 - \rho) \beta b \sigma^2 + \frac{(1 - q) \exp \left( -b^2 \sigma^2 h a_{21} \right)}{q \exp \left( -b^2 \sigma^2 h a_{11} \right) + (1 - q) \exp \left( -b^2 \sigma^2 h a_{21} \right)} \]

### 3.10 Appendix: Formulas in Continuous Time

#### 3.10.1 Equilibrium Prices under a Market-Cap Rule

At each time \( t \in [0, T] \) the index fund holds a fraction \( h \) of the index stock (the one with the higher stock price), the representative CARA investor holds an optimum portfolio that maximizes his expected utility from terminal wealth, and the stock market clears. The market is complete, as there are two tradeable stocks and two sources of uncertainty.

The investor maximizes his utility from terminal wealth:

\[
\max_{\{a_{1T}, a_{2T}\}} \quad \mathbb{E} \left[ -\exp \left( -bW_{T} \right) \right] \\
\text{subject to} \quad W_{0} \geq \mathbb{E} \left[ x_{T} W_{T} \right]
\]

\( \pi_t \) is the (unique) state-price density process at \( t \) normalized so that \( \pi_0 = 1 \). Denoting the Lagrange multiplier by \( \lambda \), the first-order condition is

\[ b \exp (-bW_{T}) - \lambda \pi_{T} = 0. \]

This determines the terminal state-price density as a function of terminal wealth. In equilibrium, the investor’s terminal wealth is equal to:

\[ W_{iT} = \left( 1 - 1_{\{D_{iT} \geq D_{iT} \}} \right) D_{iT} + \left( 1 - 1_{\{D_{iT} > D_{iT} \}} \right) D_{iT} + \Lambda_{iT} + \right. \\
\left. + \left[ W_{0} + \left( 1 - 1_{\{S_{10} \geq S_{20} \}} \right) S_{10} - \left( 1 - 1_{\{S_{20} > S_{10} \}} \right) S_{20} \right]. \]

In other words, the investor holds \( 1 - h \) of the index stock and \( h \) of the non-index stock. The remaining wealth of the investor is invested in the risk-free asset, earning a zero rate of return. Since this strategy is in fact not self-financing, the difference is accounted for by the local time of the process \( \Lambda_{iT} \) around \( S_{1T} - S_{2T} = 0 \). For now, we assume this last
stochastic term is "small," and we then write terminal wealth as a sum of a constant $c$ and the stochastic dividends of the stocks. We also define $\{1\} = \{D_{1T} \geq D_{2T}\} = \{B_{1T} \geq B_{2T}\}$ and $\{2\} = \{D_{2T} > D_{1T}\}$.

The state-price density at $t$ is then given by the conditional expectation:

$$\pi_t = E_t[\pi_T] = \frac{b \exp\left(-\frac{bc}{2}\right)}{\lambda} E_t \left[ \exp\left[-b \left( (1 - 1_{\{1\}}h) D_{1T} + (1 - 1_{\{2\}}h) D_{2T} \right) \right] \right]$$

Knowing the state-price density, we can derive the stock price from the terminal dividend:

$$S_t = \frac{E_t[\pi_T D_{T,T}]}{\pi_t}.$$ 

The difficulty arises from the indicator function which breaks down the normal distribution structure and which is also dependent on both underlying Brownian motions. We use a similar substitution as in the discrete-time case to make the index selection dependent on only one Brownian motion:

$$\tilde{B}_{1t} = B_{1t} - B_{2t},$$
$$\tilde{B}_{2t} = B_{1t} + B_{2t}.$$ 

We also distinguish between the Brownian motion realizations before and after time $t$:

$$\tilde{B}_{1tT} = \tilde{B}_{1T} - \tilde{B}_{1t} \sim N(0, 2(T-t))$$
$$\tilde{B}_{2tT} = \tilde{B}_{2T} - \tilde{B}_{2t} \sim N(0, 2(T-t)).$$

Therefore:

$$-b \left( (1 - 1_{\{1\}}h) D_{1T} + (1 - 1_{\{2\}}h) D_{2T} \right)$$
$$= -b [((2 - h) \mu T + \sigma B_{1t} + \sigma B_{2t}) - b \sigma \left( 1 - \frac{h}{2} \right) \tilde{B}_{2tT} + h b \sigma \left( 1_{\{1\}} \left( B_{1t} + \frac{\tilde{B}_{1tT}}{2} \right) + 1_{\{2\}} \left( B_{2t} - \frac{\tilde{B}_{1tT}}{2} \right) \right)] .$$

Defining

$$\tilde{B}_{1tT} = \sqrt{2(T-t)} z_1$$
$$\tilde{B}_{2tT} = \sqrt{2(T-t)} z_2$$
where \( x_1 \) and \( x_2 \) are standard normal, we can compute the expectation:

\[
E_t \left[ \exp \left[ -b \left( (1 - 1_{(1)}) D_{1T} + (1 - 1_{(2)}) D_{2T} \right) \right] \right] = \\
= \exp \left[ -b \left( (2 - h) \mu T + \sigma B_{1t} + \sigma B_{2t} \right) \right] E_t \left[ \exp \left[ \lambda \left( B_{1t} + \frac{B_{1T}}{2} \right) + 1 \left( B_{2t} - \frac{B_{2T}}{2} \right) \right] \right] \times \\
\times E_t \left[ \exp \left[ -b \sigma \left( 1 - \frac{h}{2} \right) B_{2t} \right] \right]
\]

\[
= \exp \left[ -b \left( (2 - h) \mu T + \sigma B_{1t} + \sigma B_{2t} \right) + b^2 \sigma^2 \left( 1 - \frac{h}{2} \right)^2 (T - t) \right] \times \\
\times \left[ \int_{-\infty}^{\frac{B_{2t} - B_{1t}}{\sqrt{2(T - t)}}} \exp \left[ \lambda \left( B_{2t} - \frac{B_{2t}}{2} \sqrt{2(T - t) \lambda} \right) \right] d\lambda + \right. \\
\left. \int_{\frac{B_{2t} - B_{1t}}{\sqrt{2(T - t)}}}^{\infty} \exp \left[ \lambda \left( B_{2t} + \frac{B_{2t}}{2} \sqrt{2(T - t) \lambda} \right) \right] d\lambda \right]
\]

\[
= \exp \left[ -b \left( (2 - h) \mu T + \sigma B_{1t} + \sigma B_{2t} \right) + b^2 \sigma^2 \left( 1 - \frac{h}{2} \right)^2 (T - t) + \frac{1}{4} b^2 \sigma^2 h^2 (T - t) \right] \times \\
\times \left[ \exp \left[ \lambda \left( B_{2t} \right) \right] \right] \left[ \exp \left[ \frac{1}{2} \left( \frac{B_{2t}}{\sqrt{2(T - t)}} \right)^2 \right] \right]
\]

where we defined:

\[
A_{1t} = \frac{B_{2t} - B_{1t}}{\sqrt{2(T - t)}} - b \sigma \sqrt{\frac{T - t}{2}} \\
A_{2t} = \frac{B_{2t} - B_{1t}}{\sqrt{2(T - t)}} + b \sigma \sqrt{\frac{T - t}{2}}.
\]

The expectation

\[
E_t \left[ D_{1T} \exp \left[ -b \left( (1 - 1_{(1)}) D_{1T} + (1 - 1_{(2)}) D_{2T} \right) \right] \right]
\]

is computed in the same way using the same substitutions. This gives the numerator of the stock price. Working through the algebra, we eventually obtain:

\[
S_{1t} = \mu T + \sigma B_{1t} - b \sigma^2 \left( 1 - \frac{h}{2} \right) (T - t) + \\
+ b \sigma^2 \frac{h}{2} (T - t) \frac{1}{\sqrt{2(T - t)}} \exp \left[ \frac{1}{2} \left( \frac{B_{2t}}{\sqrt{2(T - t)}} \right)^2 \right] - \frac{1}{4} b^2 \sigma^2 h^2 (T - t)
\]

\[
+ \frac{1}{2} \sigma \sqrt{\frac{T - t}{2}} \exp \left[ \frac{1}{2} \left( \frac{B_{2t}}{\sqrt{2(T - t)}} \right)^2 \right] \left[ \exp \left[ \frac{1}{2} \left( \frac{B_{2t}}{\sqrt{2(T - t)}} \right)^2 \right] \right]
\]

\[
+ \frac{1}{2} \sigma \sqrt{\frac{T - t}{2}} \exp \left[ \frac{1}{2} \left( \frac{B_{2t}}{\sqrt{2(T - t)}} \right)^2 \right] \left[ \exp \left[ \frac{1}{2} \left( \frac{B_{2t}}{\sqrt{2(T - t)}} \right)^2 \right] \right]
\]

\[
\times \exp \left[ \frac{1}{2} \left( \frac{B_{2t}}{\sqrt{2(T - t)}} \right)^2 \right]
\]

3.10.2 Certainty Equivalent

To compute the instantaneous certainty equivalent in continuous time, we need to define the value function \( J_t (W_t) \):

\[
J_t (W_t) = \max_{\{a_{1t}, a_{2t}\}} \left[ \begin{array}{c} E_t [\exp (-b W_t)] \\
\text{subject to} : W_t \geq \frac{E_t [\pi T W_T]}{\pi_t} \end{array} \right]
\]

150
The value function gives the expected utility of the investor conditional on all information available at time \( t \). We know it is equal to:

\[
J_t(W_t) = -\exp\left[-b \left(W_t - \left(1 - \mathbb{1}_{S_{1T} \geq S_{2T}} \right) S_{1T} - \left(1 - \mathbb{1}_{S_{2T} > S_{1T}} \right) S_{2T}\right)\right] \times \\
\times \mathbb{E}_t \left[\exp\left[-b \left(\left(1 - \mathbb{1}_{1T} \right) D_{1T} + \left(1 - \mathbb{1}_{2T} \right) D_{2T}\right)\right]\right].
\]

Earlier we already computed the analytical expression for this expectation. We define the instantaneous certainty equivalent \( W_{CE,t} \) as follows:

\[
\mathbb{E}_t [J_{t+dt}(W_t + W_{CE,t} dt)] = \mathbb{E}_t [J_{t+dt}(W_t)]
\]

where \( W_{t+dt} \) is wealth at time \( t + dt \) under the investment policy that we are evaluating. With exponential utility, we can write:

\[
J_{t+dt}(W_t + W_{CE,t} dt) = \exp\left[-b \left(W_t + W_{CE,t} dt\right)\right] J_{t+dt}(0)
\]

\[
J_{t+dt}(W_{t+dt}) = \exp\left[-b \left(W_t + dt\right)\right] J_{t+dt}(0).
\]

Rearranging the equation, we obtain:

\[
W_{CE,t} dt = -\frac{1}{b} \log \left[\frac{\mathbb{E}_t [\exp\left(-bdW_t\right) J_{t+dt}(0)]}{\mathbb{E}_t [J_{t+dt}(0)]}\right].
\]

Writing out the differentials separately for the terms under the expectation in the numerator:

\[
\exp(-bdW_t) = 1 - bdW_t + \frac{1}{2} b^2 (dW_t)^2
\]

\[
J_{t+dt}(0) = J_t(0) + \frac{\partial J_t(0)}{\partial B_{1t}} dB_{1t} + \frac{\partial J_t(0)}{\partial B_{2t}} dB_{2t} + \left[\frac{1}{2} \frac{\partial^2 J_t(0)}{\partial B_{1t}^2} + \frac{1}{2} \frac{\partial^2 J_t(0)}{\partial B_{2t}^2} + \frac{\partial J_t(0)}{\partial t}\right] dt
\]

Hence, we need to compute a few partial derivatives of \( J_t(0) \), which we can do, although the expressions are long.

Let \( \mu_{it} \) and \( \sigma_{it} \) be the coefficients for the price of stock \( i \) in the expression \( dS_{it} = \mu_{it} dt + \sigma_{it} dB_{1t} + \sigma_{it} dB_{2t} \).

Denoting \( y_{it} \) as the number of shares of stock \( i \) in the portfolio that we are evaluating, the wealth dynamics under this policy are given by

\[
dW_t = (y_{i1t} \mu_{1t} + y_{i2t} \mu_{2t}) dt + (y_{i1t} \sigma_{11t} + y_{i2t} \sigma_{21t}) dB_{1t} + (y_{i1t} \sigma_{12t} + y_{i2t} \sigma_{22t}) dB_{2t}.
\]

After some algebra, we find the instantaneous certainty equivalent:

\[
W_{CE,t} = (y_{i1t} \mu_{1t} + y_{i2t} \mu_{2t}) - \frac{1}{2} b \left[(y_{i1t} \sigma_{11t} + y_{i2t} \sigma_{21t})^2 + (y_{i1t} \sigma_{12t} + y_{i2t} \sigma_{22t})^2\right] +
\]

\[
(y_{i1t} \sigma_{11t} + y_{i2t} \sigma_{21t}) \frac{1}{J_t(0)} \frac{\partial J_t(0)}{\partial B_{1t}} + (y_{i1t} \sigma_{12t} + y_{i2t} \sigma_{22t}) \frac{1}{J_t(0)} \frac{\partial J_t(0)}{\partial B_{2t}}.
\]

The first terms are the expected dollar return minus the variance of this return, exactly as in a one-period CARA-normal setting. The last two terms incorporate the effect of changing future investment opportunities on the investor's expected utility. Since the value function \( J_t(0) \) depends on the optimal portfolio holdings (i.e. the portfolio of the non-
index investor) which are discontinuous at the index cutoff, the partial derivatives of $J_t(0)$ with respect to $B_{1t}$ and $B_{2t}$ are also discontinuous. This creates a discontinuity in the certainty equivalent for a given stock even though its drift and volatility are smooth around the index cutoff.
3.11 Figures

Equilibrium under the Market-Cap Rule at $t = 1$

Price of Stock 1

Price of Stock 2

Expected return of stock 1

Expected return of stock 1

Variance of return of stock 1

Cross-correlation of returns
The Effect of Risk Aversion
Equilibrium under an Exogenous Index, $q = 0.5$

Price of Stock 1

Expected return of stock 1

Variance of return of stock 1

Cross-correlation of returns
Equilibrium under an Exogenous Index, $q = 1$
Certainty Equivalents between $t = 1$ and $t = 2$ under Different Index Rules

Stock 2, market-cap rule

Index stock, market-cap rule

Index stock, $q=0.5$

Index stock, $q=0.5$ vs. market-cap

Index stock, $q=1$ vs. $q=0.5$

Investor's portfolio, market-cap rule

157
Continuous-Time, Market-Cap Rule

Price of stock 1

Price of stock 2

Drift of stock 1

Drift of stock 2

Diffusion 1 of stock 1

Diffusion 2 of stock 1
Continuous-Time, Market-Cap Rule

Volatility of stock 1

Cross-correlation

Certainty equivalent stock 2

Certainty equivalent index stock

Certainty equivalent non-index stock

Certainty equivalent investor's portfolio