What Makes Demand Curves for Stocks Slope Down?

Antti Petajisto*
Sloan School of Management
Massachusetts Institute of Technology
50 Memorial Drive, E52-442
Cambridge, MA 02142
antti@mit.edu

Comments solicited
January 16, 2003

Abstract

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 build a model to demonstrate that when investors channel a large share of their wealth through financial intermediaries, this link between the pricing of market risk and idiosyncratic risk does not have to exist. In equilibrium, the fee charged by active money managers can entirely determine the cross-sectional pricing of stocks, while the risk aversion of the end investors still determines the aggregate market risk premium. Contradicting any representative agent models, this allows demand curves for stocks to be steep enough to have economic significance. Consequently, the market prices stocks only within some bounds of their fundamental values, and prices can fluctuate within these bounds even for noninformational reasons. This could account for several empirically observed puzzles such as the S&P 500 index premium.

*I wish to thank Steve Ross, Dimitri Vayanos, Tuomo Vuolteenaho, and Jiang Wang for valuable discussions, as well as Randy Cohen, Josh Coval, Jonathan Lewellen, Andrew Lo, Oguzhan Ozbas, Anna Pavlova, Andrei Shleifer, and Finance Seminar and Finance Lunch participants at MIT for comments. I am also grateful to Frank Russell Co. for providing data for this study, and to the Emil Aaltonen Foundation and the Finnish Cultural Foundation for financial support.
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 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 return of about +14% for the 50 index additions and about -18% for the 22 deletions (Figure 2).2 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 segment-

---

1 The Wall Street Journal, 7/11/02.
2 Author’s calculations.
tation. 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 paper, our goal is to provide a theoretical explanation for downward-sloping
demand curves that satisfies two important criteria: First, it is generally applicable, po-
tentially 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.3 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, only professional money managers have information about the fundamen-
tals of individual stocks. End investors can invest in individual stocks only through these
professional active managers (stock pickers) who charge a fee for their services. 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 fee for active management.

3Author’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 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.

However, 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? It is due to the active manager’s fee, but why can the two not contract around the issue? 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 fixed cost endogenously gives rise to institutions which actually make stock prices more efficient, i.e. closer to the CAPM benchmark.

The paper 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 explains our tests for Russell 2000 and S&P 500 index changes and provides some empirical evidence consistent with our theoretical predictions. 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 almost all tables and pictures.

2 The Puzzle: Theory and Empirical Evidence

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 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 \( \tilde{x}_i = a_i + b_i \tilde{y} + \tilde{e}_i \). 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 \( \tilde{e}_i \sim N(0, \sigma_{e_i}^2) \). \(^4\) \( a_i \) and \( b_i \) are stock-specific constants. The return on the risk-free asset

\(^4\)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.
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:

$$
\max_{\{\theta_i\}} E \left[ -\exp \left( -\gamma \bar{W} \right) \right] \\
\text{s.t. } \bar{W} = W_0 + \sum_{i=1}^{N_S} \theta_i (\bar{x}_i - P_i).
$$

(1)

We calculate the first-order conditions with respect to $\theta_i$, taking the market variance $\sigma^2_m$ 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:

$$
P_i = a_i - \gamma \underbrace{\sigma^2_m \left( \sum_{j \neq i} u_j b_j \right)}_{\text{depends on systematic risk } b_i} + \underbrace{(\sigma^2_m b_i^2 + \sigma^2_{e_i}) u_i}_{\text{depends on supply } u_i}.
$$

(2)

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 $\sigma^2_{e_i} = 900$ for all stocks and $\sigma^2_m = 0.04$ 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 the section on CRRA utility later in the paper.}

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 Drunen (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 (Morek 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 14% cumulative price premium for index additions while the demand shock by mechanical indexers was approximately 10% of the shares outstanding of each stock.6

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.

### 2.4 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 paper to conduct an exhaustive investigation of each hypothesis. Hence, we cannot definitively rule out any of the hypotheses, but we can

---

6 The index premium based on author’s calculations; 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)).
point out some suggestive evidence about them.

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

2.4.2 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

---

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

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.

2.4.3 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. 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.

3 An Explanation with Financial Intermediaries

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.10 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?

3.2 The Model

We consider a setup similar to the one we used earlier for the CAPM calibration (Figure 3). 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.

10This 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.
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.

3.2.1 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 $\sigma^2_{\epsilon_i}$. 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.

3.2.2 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:

$$\max_{\{W_a, W_p\}} E \left[ -\exp \left( -\gamma_e \bar{W}_1 \right) \right]$$

s.t. $\bar{W}_1 = W_0 + W_a \bar{R}_a + W_p \bar{R}_m$,  \hspace{1cm} (3)

where $\bar{R}_a$ and $\bar{R}_m$ are the net 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 net return on stock $i$ as $\bar{R}_i$ and the price of the market portfolio as $P_m$, we can write the portfolio returns as

$$\bar{R}_a = \left( \sum_{i=1}^{N_S} v_i \bar{R}_i \right) - f_a \tag{4}$$

$$\bar{R}_m = \frac{1}{P_m} \sum_{i=1}^{N_S} P_i \bar{R}_i \tag{5}$$

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 $\bar{R}_a = \alpha_a + \beta_a \bar{R}_m + \bar{\epsilon}_a$ where $\beta_a$ is the market beta of the portfolio and $\bar{\epsilon}_a \sim N(0, \sigma^2_\epsilon)$. Then the after-fees abnormal
return $\alpha_a$ and the idiosyncratic variance $\sigma_a^2$ of the manager’s portfolio are given by:

$$\alpha_a = \sum_{i=1}^{N_S} v_i \alpha_i - f_a$$

(6)

$$\sigma_a^2 = \sum_{i=1}^{N_S} v_i^2 \sigma_i^2$$

(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:

$$W_a^* = \frac{E[R_a] - \beta_a \eta}{\gamma_e \sigma_a^2} = \frac{\alpha_a}{\gamma_e \sigma_a^2}$$

(8)

$$W_p^* = \frac{E[R_m] - \beta_a W_a^*}{\gamma_e \sigma_m^2} = \frac{\eta}{\gamma_e \sigma_m^2} - \beta_a W_a^*,$$

(9)

where $\eta$ denotes the market risk premium.

The value function of the investor can be defined as the solution to the maximization problem (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:

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

(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).

### 3.2.3 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 repack, any nonlinear pricing (including nonlinear fees) would represent an arbitrage opportunity.

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.\footnote{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 therefore set the cost of short positions equal to zero, so the cost of the portfolio is determined by its long positions only:

\[
\sum_{v_i > 0} v_i = 1. \tag{11}
\]

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

\[
f \sum_{i=1}^{N_S} |W_a v_i| = f W_a \sum_{i=1}^{N_S} |v_i|, \tag{12}
\]

which translates to a fee of

\[
f_a = f \sum_{i=1}^{N_S} |v_i| \tag{13}
\]
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, which allows the manager to learn the stock-specific parameters \( a_i, b_i, \) and \( \sigma_{\varepsilon_i}^2 \) and then 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 (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:

\[
\max_{\{v_i\}, f} \frac{\alpha_a}{\sigma_a} \frac{\sum_{i=1}^{N_S} v_i \alpha_i - f \sum_{i=1}^{N_S} |v_i|}{\sqrt{\sum_{i=1}^{N_S} v_i^2 \sigma_i^2}}
\text{s.t.} \ f W^*_a \sum_{i=1}^{N_S} |v_i| \geq C.
\]  

(14)

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

\[
\max_{\{v_i\}, f} \frac{\sum_{i=1}^{N_S} v_i \alpha_i - f \sum_{i=1}^{N_S} |v_i|}{\sum_{i=1}^{N_S} v_i^2 \sigma_i^2}
\text{s.t.} \ f \sum_{i=1}^{N_S} v_i \gamma e \sum_{i=1}^{N_S} |v_i| \sum_{i=1}^{N_S} |v_i| \geq C.
\]  

(15)

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

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

(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 (14) confirms that our proportional fee does not change the optimal portfolio weights chosen by the manager.\(^{13}\)

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.

\(^{13}\)Section 5.1 further discusses the optimality of this contract.
of interest in this paper, 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

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

where we defined the “effective risk aversion” of the active manager as

\[
\gamma = \frac{1}{W_a} \sum_{\alpha_j > 0} \frac{\alpha_j}{\sigma_j^2}.
\]

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.

### 3.2.4 Equilibrium

There are three groups of investors holding stock \( i \): First, the passive manager holds the same fraction \( u_p = \frac{W_a}{P_m} \) 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:

\[
u_p + u_{in} + u_i = 1.
\]

The equilibrium price of the stock will be

\[
P_i = \underbrace{a_i}_{\text{expected payoff}} - \underbrace{b_i \eta}_{\text{discount for market risk}} - \underbrace{\gamma \sigma_u^2 u_i}_{\text{deviation from CAPM}}.
\]

This yields an alpha of

\[
\alpha_i = \frac{\gamma \sigma_u^2}{P_i} u_i.
\]

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:

$$\eta = \frac{\gamma e \sigma^2_M}{N_S a - \gamma e \sigma^2_M}$$  \hspace{1cm} (22)

$$W_p = \frac{N_S a - \gamma e \sigma^2_M}{2}$$  \hspace{1cm} (23)

$$W_a = \frac{N_S \sigma_u}{\sigma^2_M} \left[ \sqrt{\frac{2}{\pi}} (a - b\eta) - \gamma e \sigma^2_u \right]$$  \hspace{1cm} (24)

$$f = \frac{C}{\sqrt{\frac{2}{\pi}} (a - b\eta) N_S \sigma_u}$$  \hspace{1cm} (25)

$$\gamma = \gamma e + \sqrt{\frac{2}{\pi}} \left( \frac{a - b\eta}{\sigma^2_u} \right) f.$$  \hspace{1cm} (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.

### 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 $\gamma_e = 1.5625 \times 10^{-5}$ (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), $\sigma^2_e = 900$ (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:

$$\gamma = \gamma_e + \sqrt{\frac{2}{\pi}} \left( \frac{a - b\eta}{\sigma^2_u} \right) f.$$  \hspace{1cm} (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 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 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 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.

<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>

Table 2: The effect of the management fee; five-year horizon.

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 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 paper 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.

3.4 Empirical Implications

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

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

The price of a stock is given by its CAPM price \((a_i - b_i \eta)\) minus a deviation \((\gamma \sigma_{e_i}^2 u_i)\) due to idiosyncratic risk. As the equilibrium holdings \((u_i)\) of the active manager change, the price impact is given by the dollar variance \((\sigma_{e_i}^2)\) 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

\[ \frac{dQ_i}{Q_i} = \frac{dP_i}{P_i} = P_i \frac{du_i}{dP_i} = -\frac{P_i}{\gamma \sigma_{e_i}^2}. \]

**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 noise trader holdings are inversely correlated with fundamental news.*

### 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.

Instead, the link of Implication 1 between idiosyncratic risk and the demand curve slope provides an opportunity for a relatively straightforward test. We perform 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 investigate index additions rather than deletions because the deletions from the S&P 500 represent a much smaller sample than the additions.\textsuperscript{14} In the future we plan to extend the tests to deletions for both indices.

#### 4.1 The Russell 2000

##### 4.1.1 Background

Each year, Frank Russell Co. sorts all publicly listed U.S. firms based on the market capitalizations of their public floats after the close of trading on May 31. The largest

\textsuperscript{14}This is due to mergers between two index firms.
1,000 stocks form the Russell 1000 index and the largest 3,000 stocks form the Russell 3000 index. The Russell 2000 index consists of all Russell 3000 stocks minus all Russell 1000 stocks. The Russell 2000 thus measures the performance of small-cap stocks, and for that purpose it is still the most commonly quoted index. The new index composition becomes effective on July 1, and that will generally not change for the next 12 months unless the stock is delisted from its exchange. The index weight of a stock is determined by the market value of its public float as defined by Russell.

The index funds mechanically tracking the Russell 2000 generally update their portfolios on the last trading day of June in order to minimize their tracking error. The index composition is announced much earlier in June though, and any active market participant can infer the index composition based on market prices on May 31. Since the index selection rule is entirely mechanical, it is in fact possible for market participants to have a good idea of the future index composition even before May 31. However, historically most of the price impact associated with index changes has taken place in June, so in order to minimize the effect of noise and also not to create any forward-looking bias, we focus on abnormal returns in June.

Stocks can be added to the Russell 2000 from below or above, i.e. if they rise above the lower cutoff (recently around $150 million) or fall below the upper cutoff (recently around $1.5 billion). Stocks crossing the upper cutoff experience a demand shock due to the different fractions held by Russell 1000 and Russell 2000 indexers. We focus on stocks added from below because it represents a cleaner experiment.

Our data for the Russell indices is obtained from Frank Russell Co. Prior to 1987, the Russell indices were reconstituted quarterly, so we only use data from 1987 to 2000.

4.1.2 Methodology

We want to test whether the idiosyncratic risk of a stock affects the price impact around index addition. 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 of the stock.

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. We define idiosyncratic risk as the root mean squared error of this regression. 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

\footnote{http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html}

\footnote{We also ran all tests with the market model and obtained very similar results.
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 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.

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.

4.1.3 Results

Table 1 shows the equal-weighted cumulative abnormal return for the Russell 2000 additions that satisfy our data requirement. The number of event stocks each year is now about ten times as large as for the S&P 500. The recent years have generally seen a greater index addition premium than the years in the past.

The univariate Fama-MacBeth regression produces a significant $t$-statistic of 2.15 for the coefficient of idiosyncratic risk (Table 2). The $t$-statistic drops to 1.60 when we add log market equity into the regression. However, market equity does not appear to be statistically significant either in the univariate or bivariate regression. 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 0.1 in annual idiosyncratic volatility would increase the price impact of Russell 2000 addition by 0.3%, or about 10% of the average price impact over this period. 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.

4.2 The S&P 500

4.2.1 Background

Unlike the Russell indices, the S&P 500 is updated at apparently random times throughout the year, with no significant concentrations around any particular days. The criteria for index membership include market capitalization, liquidity, size of the public float, and industry representation, but in the end both the firms involved and the exact timing of the changes are decided behind closed doors and somewhat subjectively by the S&P index membership committee.

Index changes are typically announced five trading days before the effective day. Most of the associated price jump occurs immediately after the announcement, but surprisingly a nontrivial part of the price effect takes place gradually during the intervening days between the announcement and effective day. Prior to October 1989, the index changes became effective immediately after the announcement. In order not to mix these potentially different regimes, we restrict ourselves to data from January 1990 to December 2000.

In fact, 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. We want to verify this result with more recent data, especially as the fraction of mechanical indexers has grown so much since the 1976-1989 period, and with our empirical methodology. This also allows us to link the S&P 500 results to the Russell 2000 results.

4.2.2 Methodology

As before, we want to test whether idiosyncratic risk is related to the price impact around index addition. For the event window, we choose the five trading days leading up to the official index change. We do not have the actual announcement dates of each addition, but our approximation does match the event dates for a typical index addition, and elsewhere it will work against us.

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 at least one 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. 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, using White’s heteroskedasticity-consistent standard errors. We also plug in a dummy variable for each year to account for the increasing time trend in the index premium.

4.2.3 Results

The regression results (except for the dummy variables) are shown in Table 5. Idiosyncratic risk turns out to be statistically significant both in the univariate regression ($t = 2.42$) and in the bivariate regression ($t = 2.06$). Its coefficient is about 0.1, meaning that an increase in annualized idiosyncratic volatility of 0.1 would increase the abnormal return around index addition by 1 percentage point.

5 Interpretations and Further Discussion

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.\textsuperscript{17}

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 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.

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 received 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 ab-

\textsuperscript{17}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.
sence 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.\footnote{See Berk (1995) for a discussion of similar issues. We also build a model where we explicitly deal with wide dispersion in operating sizes of firms.} Even the variation in price-to-book or 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 the table before). 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.

5.3 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. Needless to say, this means 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.

19 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 of the transaction cost; otherwise they would take extreme positive and negative positions in individual stocks.
5.4 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 $[\frac{1}{17}, \frac{7}{60}]$. 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.

5.5 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?

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.
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 paper 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%. We confirm the robustness of our results in a richer and arguably more realistic model, and the effects only become stronger.

It is relatively difficult to find clean empirical tests for the broad implications of the model, including the link between the fee for active management and the slopes of the demand curves for stocks. However, the model also predicts a cross-sectional link between the slope and idiosyncratic risk, and this we confirm using data on both the Russell 2000 and S&P 500 index changes.

We believe this paper 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 produce this result.
References


A More Elaborate Model

A.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 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. 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’Avolio (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%. 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 5.5, 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.

A.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 $\tilde{X}_i = a_i + b_i \tilde{Y} + \tilde{\epsilon}_i$. 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.

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.

---

22 Data from the exchanges, published monthly by the Wall Street Journal.
A.2.1 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 $\sigma^2_{\epsilon_i} = P_i^2 \sigma^2_{\epsilon}$, so that each stock will always have a market beta $\beta_i = 1$ and a fixed return variance of $\sigma^2_{\epsilon}$. These assumptions have a negligible effect on our numerical results but they do make our equations more convenient and intuitive.

A.2.2 End Investors

The representative end investor’s problem is again

$$\max_{\{W_a, W_p\}} E \left[ -\exp \left( -\gamma \epsilon \bar{W}_1 \right) \right]$$

subject to

$$\bar{W}_1 = W_0 + W_a \bar{R}_a + W_p \bar{R}_m,$$

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

$$W_a = \frac{E \left[ \bar{R}_a \right] - \beta_a \eta}{\gamma_a \sigma^2_{\epsilon_a}} = \frac{\alpha_a}{\gamma_a \sigma^2_{\epsilon_a}}$$

$$W_p = \frac{E \left[ \bar{R}_m \right] - \beta_a W_a}{\gamma_m \sigma^2_{\epsilon_m}} = \frac{\eta}{\gamma_m \sigma^2_{\epsilon_m}} - \beta_a W_a.$$  

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.

A.2.3 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 $a_{ik} \sim U(a_i - \Delta a, a_i + \Delta a)$. 

35
For 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 \( \tilde{z}_i = a_i + \tilde{e}_i \) and prices \( P_{e_i} = P_i (1 + \eta) \). This determines the dollar demand of the active manager \( k \) for stock \( i \):

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

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

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

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), 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.

### A.2.4 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 \( u_p = \frac{W_p}{\gamma \sigma_i^2} \), the noise traders who hold a randomly chosen supply of \( u_{in} \sim U(0, \Delta u) \), and the active managers who hold the remaining supply which we denote as \( u_i \). Market clearing then requires that

\[
u_p + u_{in} + u_i = 1
\]

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

We assume there is a continuum of managers with a measure of \( N_i \) who know about stock \( i \). Their total dollar demand for stock \( i \) is then

\[
W_i = \begin{cases} 
\int_{a = a_i + \Delta a_i}^{a_i} \frac{1}{\gamma \sigma_i^2} \left[ \frac{a}{P_i} - (1 + \eta) \right] N_i \frac{da}{2 \Delta a_i} & \text{if } P_i \geq \frac{a_i - \Delta a_i}{(1 + \eta)} \\
\int_{a = a_i - \Delta a_i}^{a_i} \frac{1}{\gamma \sigma_i^2} \left[ \frac{a}{P_i} - (1 + \eta) \right] N_i \frac{da}{2 \Delta a_i} & \text{if } P_i < \frac{a_i - \Delta a_i}{(1 + \eta)}.
\end{cases}
\]

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
\[
P_i = \frac{a_i + \Delta a_i}{1 + \eta + 2\sigma_i \sqrt{\frac{\Delta a_i}{N_i} u_i}} = \frac{a_i (1 + \Delta_i)}{1 + \eta + 2\sigma_i \sqrt{\frac{\Delta a_i}{N_i} u_i}}
\]  
where we defined the relative dispersion-of-beliefs parameter $\Delta_i = \frac{\Delta a_i}{N_i}$ 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
\[
\alpha_i = \frac{1}{1 + \Delta_i} \left[ 2\sigma_i \sqrt{\frac{\Delta a_i}{N_i} u_i} - \Delta_i (1 + \eta) \right].
\]

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 $\eta + 2\sigma_i \sqrt{\frac{\Delta a_i}{N_i} u_i}$ 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:
\[
\begin{align*}
\alpha_m &= 0 \quad (41) \\
W_a &= \frac{\alpha_a}{\gamma e \sigma^2} \quad (42) \\
W_p &= \frac{\eta}{\gamma e \sigma^2} - W_a \quad (43) \\
\gamma &= \frac{K}{W_a} \sum_{i=1}^M \frac{\alpha_{ik}}{\sigma^2_i} \quad (44) \\
u_{\text{min}} &= 1 - \frac{W_p}{P_m} - \Delta u \quad (45)
\end{align*}
\]

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 $\sigma_a^2$ 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 (40), 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).

### A.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 $N_i = \frac{KM}{N} = 1$, and we assume this is proportional to the expected payoff $a_i$ which implies a density $\lambda_i = \frac{1}{105}$ 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 managers. Finally, we choose the maximum dispersion of beliefs $\Delta_i$ 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 $\Delta_i$, 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), $P_i$, and $a_i$, as well as the allocations $W_u$ 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 3.

<table>
<thead>
<tr>
<th>fee</th>
<th>$\frac{W}{W_u+W_p}$</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>$1.72 \times 10^{-3}$</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 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 $\Delta_i$. 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

38
in beliefs \( \Delta \), 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.
B Figures

Figure 1: July 2002 deletion of non-U.S. firms from S&P 500.

Figure 2: S&P 500 index premium in 1999-2000.
Figure 3: The basic setup for the model.
### C Tables

#### C.1 Medians for the Benchmark Portfolios 1987-2000

<table>
<thead>
<tr>
<th>Size</th>
<th>Idiosyncratic risk</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>655</td>
<td>654</td>
<td>655</td>
<td>654</td>
<td>654</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>172</td>
<td>172</td>
<td>172</td>
<td>172</td>
<td>171</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>108</td>
<td>108</td>
<td>108</td>
<td>108</td>
<td>108</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>76</td>
<td>76</td>
<td>76</td>
<td>76</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>60</td>
<td>59</td>
<td>59</td>
<td>59</td>
<td>59</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>42</td>
<td>41</td>
<td>42</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>39</td>
<td>39</td>
<td>39</td>
<td>39</td>
<td>38</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>35</td>
<td>35</td>
<td>34</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>Big</td>
<td></td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>33</td>
</tr>
</tbody>
</table>

Table 4: Median number of stocks in each portfolio.

<table>
<thead>
<tr>
<th>Size</th>
<th>Idiosyncratic risk</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>35</td>
<td>32</td>
<td>29</td>
<td>20</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>109</td>
<td>105</td>
<td>102</td>
<td>101</td>
<td>98</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>202</td>
<td>202</td>
<td>198</td>
<td>195</td>
<td>191</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>325</td>
<td>324</td>
<td>324</td>
<td>319</td>
<td>319</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>516</td>
<td>510</td>
<td>513</td>
<td>507</td>
<td>510</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>805</td>
<td>801</td>
<td>799</td>
<td>783</td>
<td>776</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>1,223</td>
<td>1,241</td>
<td>1,228</td>
<td>1,208</td>
<td>1,229</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>2,089</td>
<td>2,073</td>
<td>2,017</td>
<td>2,026</td>
<td>1,951</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>3,695</td>
<td>3,860</td>
<td>3,931</td>
<td>3,726</td>
<td>3,593</td>
</tr>
<tr>
<td>Big</td>
<td></td>
<td>18,541</td>
<td>18,376</td>
<td>14,558</td>
<td>12,478</td>
<td>11,712</td>
</tr>
</tbody>
</table>

Table 5: Median market capitalization.
### Idiosyncratic risk (annualized)

<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 6: Median idiosyncratic risk.

### 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 7: Median share of stock-days with zero trading volume.
## C.2 Russell 2000

Table 8: Russell 2000 additions in 1987-2000. This is the average number of event stocks per year for each benchmark portfolio.

<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>0.1</td>
<td></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>
<tr>
<td>Year</td>
<td>N</td>
<td>CAR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>----</td>
<td>-----</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1987</td>
<td>234</td>
<td>0.028</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1988</td>
<td>176</td>
<td>0.055</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1989</td>
<td>102</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>283</td>
<td>0.026</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td>381</td>
<td>0.024</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td>423</td>
<td>0.020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td>324</td>
<td>0.016</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td>423</td>
<td>0.003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>334</td>
<td>0.041</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>380</td>
<td>0.028</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>407</td>
<td>0.060</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>389</td>
<td>0.044</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>346</td>
<td>0.087</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>467</td>
<td>0.091</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9: Abnormal returns for Russell 2000 additions. These are the value-weighted cumulative abnormal returns (i.e. each stock’s return over its benchmark portfolio return) for all event stocks June 1 through June 30 of the corresponding year.

<table>
<thead>
<tr>
<th>intercept</th>
<th>sigma</th>
<th>$\log_{10}(ME)$</th>
<th>$t_{intercept}$</th>
<th>$t_{sigma}$</th>
<th>$t_{\log_{10}(ME)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.018</td>
<td>0.033</td>
<td></td>
<td>1.701</td>
<td>2.155</td>
<td></td>
</tr>
<tr>
<td>0.077</td>
<td>-0.025</td>
<td></td>
<td>1.513</td>
<td>-1.097</td>
<td></td>
</tr>
<tr>
<td>0.068</td>
<td>0.023</td>
<td>-0.026</td>
<td>1.361</td>
<td>1.600</td>
<td>-1.220</td>
</tr>
</tbody>
</table>

Table 10: Fama-MacBeth regression results for Russell 2000 additions. These are the gamma estimates and their t-statistics over 14 annual cross-sections, 1987-2000.
C.3 S&P 500

<table>
<thead>
<tr>
<th>Number of stocks</th>
<th>Idiosyncratic risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>Low</td>
</tr>
<tr>
<td>Small</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Big</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 11: S&P 500 additions in 1990-2000. This is the number of event stocks for each benchmark portfolio.

<table>
<thead>
<tr>
<th>intercept</th>
<th>sigma</th>
<th>$log_{10}(ME)$</th>
<th>$t_{intercept}$</th>
<th>$t_{sigma}$</th>
<th>$t_{log_{10}(ME)}$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.011</td>
<td>0.103</td>
<td></td>
<td>0.479</td>
<td>2.418</td>
<td></td>
<td>0.059</td>
</tr>
<tr>
<td>-0.134</td>
<td></td>
<td></td>
<td>-1.112</td>
<td>1.673</td>
<td></td>
<td>0.059</td>
</tr>
<tr>
<td>-0.166</td>
<td>0.092</td>
<td>0.047</td>
<td>-1.390</td>
<td>2.063</td>
<td>1.478</td>
<td>0.082</td>
</tr>
</tbody>
</table>

Table 12: Abnormal returns for S&P 500 additions in 1990-2000 regressed on idiosyncratic risk, log market equity, and dummy variables for each year (not shown in the table). The t-statistics are based on White’s heteroskedasticity-consistent standard errors.
D Derivations of Formulas

D.1 CAPM Benchmark

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

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

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

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

$$\max_{\{\theta_i\}} \sum_{i=1}^{N_S} \theta_i (a_i - P_i) - \frac{1}{2} \gamma \left( \sum_{i=1}^{N_S} \theta_i b_i \right)^2 \sigma_m^2 - \frac{1}{2} \gamma \sum_{i=1}^{N_S} \theta_i^2 \sigma_{e_i}^2.$$

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

$$a_i - P_i - \gamma \left( \sum_{j=1}^{N_S} \theta_j b_j \right) b_i \sigma_m^2 - \gamma \sigma_{e_i}^2 \theta_i = 0.$$  

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

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

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

D.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:

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

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

We write the Lagrangian of this problem as

$$\sum_{j=1}^{N_S} v_j \alpha_j - f \sum_{j=1}^{N_S} |v_j| + \lambda \left[ f \sum_{j=1}^{N_S} v_j \alpha_j - f \sum_{j=1}^{N_S} |v_j| \sum_{j=1}^{N_S} |v_j| - C \right].$$

The first-order conditions with respect to \( v_j \) and \( f \) yield:

$$\frac{[\alpha_j - f(v_j)](\Sigma_{j=1}^{N_S} v_j \sigma_j)}{\Sigma_{j=1}^{N_S} v_j^2 \sigma_j^2} - \frac{\Sigma_{j=1}^{N_S} v_j \sigma_j - f \Sigma_{j=1}^{N_S} |v_j|}{\Sigma_{j=1}^{N_S} v_j^2 \sigma_j^2} - \frac{1}{2} \Sigma_{j=1}^{N_S} v_j^2 \sigma_j^2 - \frac{1}{2} \Sigma_{j=1}^{N_S} v_j \sigma_j - f \Sigma_{j=1}^{N_S} |v_j| \Sigma_{j=1}^{N_S} v_j^2 \sigma_j^2 - \frac{1}{2} \Sigma_{j=1}^{N_S} v_j \sigma_j - f \Sigma_{j=1}^{N_S} |v_j| \Sigma_{j=1}^{N_S} v_j^2 \sigma_j^2 - \frac{1}{2} \Sigma_{j=1}^{N_S} v_j \sigma_j - f \Sigma_{j=1}^{N_S} |v_j| \Sigma_{j=1}^{N_S} v_j^2 \sigma_j^2 - \frac{1}{2} \Sigma_{j=1}^{N_S} v_j \sigma_j - f \Sigma_{j=1}^{N_S} |v_j| \Sigma_{j=1}^{N_S} v_j^2 \sigma_j^2 - \frac{1}{2} \Sigma_{j=1}^{N_S} v_j \sigma_j - f \Sigma_{j=1}^{N_S} |v_j| \Sigma_{j=1}^{N_S} v_j^2 \sigma_j^2 - \frac{1}{2} \Sigma_{j=1}^{N_S} v_j \sigma_j - f \Sigma_{j=1}^{N_S} |v_j| \Sigma_{j=1}^{N_S} v_j^2 \sigma_j^2 - \frac{1}{2} \Sigma_{j=1}^{N_S} v_j \sigma_j - f \Sigma_{j=1}^{N_S} |v_j| \Sigma_{j=1}^{N_S} v_j^2 \sigma_j^2$$

$$= 0$$

$$\sum_{j=1}^{N_S} v_j^2 \sigma_j^2 + \lambda \left[ \Sigma_{j=1}^{N_S} v_j \sigma_j - f \Sigma_{j=1}^{N_S} |v_j| \right]$$

$$= 0$$

47
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:

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

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

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

### D.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:

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

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

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 \( \tilde{z}_i = a_i + \tilde{e}_i \), and its price today will be \( P_{\tilde{z}_i} = P_i + b_i \eta \). We then express the active manager’s problem as:

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

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

where

\[
\tilde{R}_z = \frac{\tilde{z}_i - a_i + \tilde{e}_i}{P_{\tilde{z}_i}} - 1 = \frac{a_i + \tilde{e}_i - (P_i + b_i \eta)}{P_i + b_i \eta}
\]

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

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

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

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

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

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

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
\[ W_m = \frac{\eta}{\gamma \sigma_m^2} \] (60)
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 (50) 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.

D.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:
\[
\begin{align*}
\frac{u_i P_i}{W_i} &= \frac{\alpha_i}{\gamma \sigma_i^2} = \frac{(a_i - b_i \eta - P_i) P_i}{\gamma \sigma_i^2}, \\
P_i &= a_i - b_i \eta - \gamma \sigma_i^2 u_i.
\end{align*}
\] (61)

The alpha of the stock is then:
\[
\alpha_i = E \left[ \frac{\bar{x}_i}{P_i} \right] - \beta_i \eta - 1 = \frac{a_i - b_i \eta - P_i}{P_i} = \frac{\gamma \sigma_i^2 u_i}{P_i}. 
\] (63)

The market portfolio has an alpha of zero by construction. Hence,
\[
\begin{align*}
\alpha_m &= \sum_{i=1}^{N_S} P_i \alpha_i = 0, \\
\sum_{i=1}^{N_S} P_i \alpha_i &= \sum_{i=1}^{N_S} P_i \frac{\gamma \sigma_i^2 u_i}{P_i} = \gamma \sigma_m^2 \sum_{i=1}^{N_S} u_i = 0, \\
\sum_{i=1}^{N_S} u_i &= 0.
\end{align*}
\] (65)

Since \( u_p + u_m + u_i = 1 \), and since the passive manager’s position \( u_p \) is constant across stocks while the noise trader’s position \( u_m \) is distributed as \( N(0, \sigma_m^2) \), the above equation implies the same distribution for the active manager’s equilibrium share holdings \( u_i \):
\[
u_i \sim N \left( 0, \sigma_i^2 \right).
\] (67)

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:
\[
u_p = \frac{W_p}{P_m} = 1.
\] (68)

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
\[
\begin{align*}
1 &= \frac{W_p}{P_m} = \left( \frac{\eta}{\gamma \sigma_m^2} \right) \frac{1}{P_m} = \left( \frac{\eta P_m^2}{\gamma \sigma_m^2} \right) \frac{1}{P_m} = \frac{\eta a_m}{\gamma \sigma_m^2 (1 + \eta)}, \\
\eta &= \frac{\gamma \sigma_m^2}{a_m - \gamma \sigma_m^2}.
\end{align*}
\] (69)

and the equilibrium allocation to the passive manager
\[
W_p = P_m = \frac{a_m}{1 + \eta} = a_m - \gamma \sigma_m^2.
\] (71)

49
To find out the allocation to the active manager, we need to find the before-fees alpha of the manager:

$$\alpha_{bf} = \frac{\sum_{u_i > 0} N_S P_i u_i \alpha_i}{\sum_{u_i > 0} P_i u_i}. \tag{72}$$

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

$$\sum_{i=1}^{N_S} P_i u_i \alpha_i = \sum_{i=1}^{N_S} \gamma \sigma_a^2 \sum_{i=1}^{N_S} u_i^2 = \gamma \sigma_a^2 N_S \sigma_u^2. \tag{73}$$

For this aggregation, we used the assumption that there is a continuum of stocks with a measure of $N_S$, so $\frac{1}{N_S} \sum_{i=1}^{N_S} u_i^2 = E [u_i^2] = \sigma_u^2$. 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

$$\sum_{u_i > 0} P_i u_i = \sum_{u_i > 0} (a_i - b_i \eta - \gamma \sigma_a^2 u_i) u_i = \gamma \sum_{u_i > 0} \sigma_a^2 u_i - \gamma \sum_{u_i > 0} \sigma_a^2 u_i^2$$

$$= \frac{N_S \sigma_u}{2} \left[ \frac{2}{\pi} (a - b \eta) - \gamma \sigma_a^2 \sigma_u \right]. \tag{74}$$

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

$$\sigma_u^2 = \frac{\sum_{i=1}^{N_S} u_i^2 \sigma_e^2}{\left( \sum_{u_i > 0} P_i u_i \right)^2} = \frac{4 \sigma_e^2}{N_S \left[ \frac{2}{\pi} (a - b \eta) - \gamma \sigma_a^2 \sigma_u \right]^2}. \tag{75}$$

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

$$f_a = \frac{\sum_{i=1}^{N_S} |W_i|}{\sum_{u_i > 0} P_i u_i} \tag{76}$$

where the numerator is

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

and thus we get

$$f_a = \frac{2 \sqrt{\frac{2}{\pi}} (a - b \eta) f}{\sqrt{\frac{2}{\pi}} (a - b \eta) - \gamma \sigma_a^2 \sigma_u}. \tag{77}$$

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

$$W_a = \frac{\alpha_{bf} - f_a}{\gamma \sigma_a^2} = \frac{N_S \left[ -\sqrt{\frac{2}{\pi}} (a - b \eta) f + \gamma \sigma_a^2 \sigma_u \right]}{2 \gamma \sigma_a^2} \left[ \sqrt{\frac{2}{\pi}} (a - b \eta) - \gamma \sigma_a^2 \sigma_u \right]. \tag{78}$$

Equating this with the cost of the manager’s portfolio (74), we obtain the simple formula for the effective risk aversion of the manager:

$$\gamma = \gamma_e + \sqrt{\frac{2}{\pi}} \left( \frac{a - b \eta}{\sigma_a^2 \sigma_u} \right) f. \tag{79}$$

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

$$C = f \sum_{i=1}^{N_S} |W_i| = \sqrt{\frac{2}{\pi}} (a - b \eta) N_S \sigma_u f \tag{80}$$

$$f = \frac{C}{\sqrt{\frac{2}{\pi}} (a - b \eta) N_S \sigma_u}. \tag{81}$$