Why Do Demand Curves for Stocks Slope Down?

Antti Petajisto*
Yale School of Management
International Center for Finance
P.O. Box 208200
New Haven, CT 06520-8200
antti.petajisto@yale.edu

May 24, 2007

Abstract

Representative agent models are inconsistent with existing empirical evidence for steep demand curves for individual stocks. This paper resolves the puzzle by proposing that stock prices are instead set by two separate classes of investors. While the market portfolio is still priced by individual investors based on their collective risk aversion, those individual investors also delegate part of their wealth to active money managers who use that capital to price stocks in the cross-section. In equilibrium the fee charged by active managers has to equal the before-fee alpha they earn; this endogenously determines the amount of active capital and the slopes of demand curves. A calibration of the model reveals that demand curves can indeed be steep enough to match the magnitude of many empirical findings, including the price effects for stocks added to (or deleted from) the S&P 500 index.

JEL classification: G12, G14, G20, D50

Keywords: Demand curves for stocks; delegated portfolio management; equilibrium mispricing; index premium

---

*I wish to thank Steve Ross, Dimitri Vayanos, Tuomo Vuolteenaho, and Jiang Wang for valuable discussions, as well as Nick Barberis, Randy Cohen, Josh Coval, John Griffin, Jonathan Ingersoll, Jonathan Lewellen, Andrew Lo, Robert Merton, Oguzhan Ozbas, Anna Pavlova, Andrei Shleifer, Matthew Spiegel, Jeremy Stein, Ivo Welch and seminar participants at Harvard University, Helsinki School of Economics, INSEAD, Massachusetts Institute of Technology, Rice University, University of Alberta, University of Michigan, Yale University, EFMA 2003 annual meeting, and NBER Universities Research Conference for comments. I am also grateful to the Emil Aaltonen Foundation and the Finnish Cultural Foundation for financial support. This paper is a revised version of Chapter 2 of my dissertation submitted to the MIT Sloan School of Management in June 2003.

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 on July 19, and they included large firms like Royal Dutch, 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. In spite of its size and publicity, this event produced a very significant price impact which showed no signs of reversal, at least in the following two months (Figure 1).

This type of evidence has led a growing empirical literature to conclude that demand curves for stocks slope down (Shleifer (1986) and Harris and Gurel (1986) are early references). In the presence of steep downward-sloping demand curves, index changes will trigger mechanical purchases and sales by index funds, which in turn can move prices. The

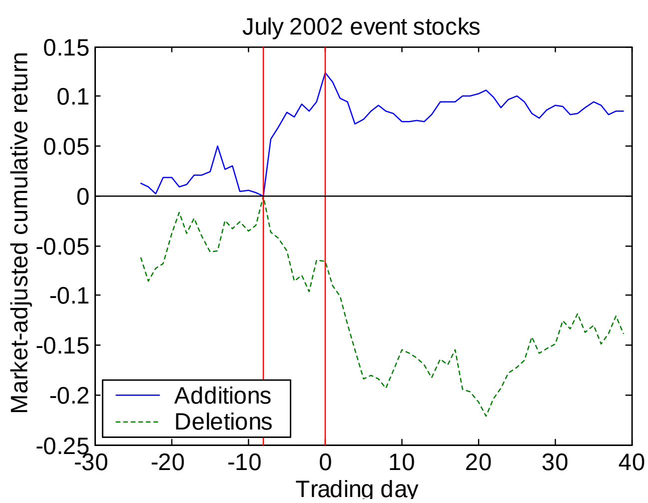

Figure 1: July 2002 replacement of seven non-U.S. firms in the S&P 500 index. The announcement occurred after the close on trading day -8, while the changes became effective at the close on trading day 0. The graph shows buy-and-hold returns on portfolios formed (initially with equal weights) on trading day -8.

[1] *Wall Street Journal*, 7/11/02.

typical price effect for both additions and deletions has been about 10% for the S&P 500 in recent years, and other widely tracked indexes have exhibited comparable demand elasticities (Petajisto (2004)).

However, this empirical evidence creates a fundamental puzzle: How do we reconcile the large magnitude of the price effect with asset pricing theory? In neoclassical finance, price equals expected future cash flows discounted by systematic risk, so the demand curve for a stock should be (almost) perfectly horizontal and we should observe (virtually) no price impact. Asymmetric information[2] cannot explain the significant price effects, because the puzzle here has to do with *clearly uninformed* supply shocks as illustrated by the above S&P 500 event.[3]

The limits of arbitrage literature has been suggested as a way to bridge this gap between theory and empirical work (Barberis and Thaler (2003) and Wurgler and Zhuravskaya (2002)). Mechanisms such as noise trader risk (De Long et al. (1990)) and performance-based arbitrage (Shleifer and Vishny (1997)) can indeed influence the pricing of non-diversifiable risk, but they cannot explain why investors are so reluctant to take *diversified* positions in individual stocks.

This paper first shows that existing equilibrium models underestimate the actual slopes of demand curves for stocks by several orders of magnitude. It then proposes a theoretical equilibrium model that can produce a realistically large magnitude for the slopes of demand curves, and not only for index additions and deletions but for *all* the stocks in the economy.

Despite the frequent references to indexing, this paper is about much more than that. Indexing just happens to provide a relatively clean empirical test for demand curves, but also non-index evidence points in the same direction: demand curves for individual stocks are steep *in general*. This implies nontrivial ineffiency in prices, because even completely uninformed demand shocks will move prices, which clearly contradicts existing neoclassical theory.

To illustrate the failure of traditional pricing models, consider the following CAPM calibration: The U.S. stock market capitalization at the end of 2002 was about $11 trillion, which means that collectively people invested $11 trillion in the market portfolio, perhaps

[2] Some examples are Grossman and Stiglitz (1980), Glosten and Milgrom (1985), and Kyle (1985).

[3] Denis, McConnell, Ovtchinnikov, and Yu (2003) actually find evidence that regular S&P 500 index changes (unlike the event we picked) may not be completely free of information. But index changes even for mechanical rule-based indexes such as the Russell 2000 exhibit comparable price effects.

expecting about a 5% annual risk premium and 20% annual volatility. This information allows us to back out their risk aversion. Now let us assume that the price of one stock changes slightly for noninformational reasons so that the investors suddenly perceive the stock to have an annual alpha of +1%, with idiosyncratic annual volatility of 30%. The investors should then immediately pour $1 trillion into that stock – more than three times the market capitalization of General Electric.[4] In other words, even a 1% annual alpha would be absurdly large in a CAPM setting. A representative investor who is willing to invest $11 trillion in the market portfolio should be extremely aggressive when any mispricings occur for individual stocks. More generally, this calibration shows that no model with a single representative investor can simultaneously generate realistic demand curves for individual stocks and a plausible market risk premium.

Building on this key insight, we argue that demand curves seem too steep only when we assume the *same* group of investors prices both the market portfolio and the cross-section of individual stocks. The puzzle disappears if we separate these roles: in particular, we let the collective actions of individual investors determine the pricing of the aggregate market, while the cross-sectional pricing is independently determined by the actions of professional money managers.

We present our story in a simple model similar to the CAPM setting. There are only two differences: First, we assume there is a fixed cost for actively managing a stock portfolio; if one does not pay the cost, one can only invest in the market portfolio. We interpret this as costly information acquisition; if one does not know about individual stocks, one's best bet is the market portfolio. Second, we assume the fixed cost is paid through a financial institution as a proportional fee.

Hence, the model introduces a layer of professional money managers between stocks and individual investors. Active managers act as stock pickers, using all their delegated wealth to take positions in individual stocks, and they charge a fee for their services. Individual investors then choose their optimal allocation of wealth between an actively managed portfolio, a passively managed market portfolio (with zero fee), and a risk-free asset. We refer to the individual investors as "end investors" because they are the ones that own all the

[4] The optimal dollar investment for a CARA or CRRA investor is proportional to $\frac{\mu}{\sigma^2}$. This is $\frac{0.05}{0.2^2} = 1.25$ for the market portfolio and $\frac{0.01}{0.3^2} = 0.11$ for the idiosyncratic gamble, producing a dollar investment of $\frac{0.11}{1.25} \times \$11 = \$1$ trillion in the idiosyncratic gamble.

wealth and derive utility from it, in contrast to the money managers who make investment decisions but do not actually own the wealth they invest.[5] The remaining supply of each stock is passively held by exogenous noise traders – without this group, even the active managers would just have to hold the market portfolio. We do not consider agency issues, so the only friction we introduce relative to the CAPM is the fixed cost which generates a fee for active management.

We find that this delegation of portfolio management completely changes the cross-sectional pricing of stocks. Now the slopes of demand curves are no longer determined by end investors' risk aversion – instead they depend on the wealth allocated to active managers, which in turn depends on the fee charged by the active managers. If the fee is 1.5% per year, then the typical stock will be "mispriced" so that it will have an alpha of either $+0.75\%$ or -0.75% per year in equilibrium (thus adding up to a 1.5% alpha in a long-short portfolio). If on average such mispricings are corrected slowly over several years,[6] then these annual alphas will be capitalized into much greater variation in stock prices today. E.g. an annual alpha of $+1\%$, fully corrected over 5 years, means a stock is underpriced by 5%. Thus the initial mispricings created by the management fee are further magnified by their slow expected convergence to fundamental values, and this allows economically large fluctuation in stock prices today. For comparison, if we set the active managers' fee to zero, pricing collapses to the traditional CAPM benchmark where annual alphas are always well within 1 bp from zero.

Yet the presence of institutions does not create any friction in the model – the true source of friction is the underlying fixed cost. The institutions actually mitigate the effect of the fixed cost and produce the *flattest* possible demand curves, because they allow the risk of active trading to be shared among all investors in the economy. Consistent with the predictions of functional and structural finance (Merton and Bodie (2002)), our institutional structure can be viewed as an endogenous outcome that minimizes price distortions due to the underlying market friction.[7]

Empirical evidence appears generally consistent with our equilibrium. Active fund managers do have some stock-selection ability (e.g., Wermers (2000) and Daniel et al. (1997)),

---

[5] The "end investors" could also be institutions such as university endowments. Note that in standard models all investors are end investors, i.e., there are no intermediaries making investment decisions.

[6] For empirical evidence on the slow correction of mispricings, see section 3.3.4.

[7] The formal analysis behind this is available from the author.

especially if they concentrate on relatively few industries (Kacperczyk, Sialm, and Zheng (2004)) or if they are small (Chen, Hong, Huang, and Kubik (2004)), but once their fees and expenses are taken into account, their alphas fall back to approximately zero. For our pricing results it is crucial that active managers indeed earn positive before-fee alphas, but whether their alphas exactly cover their fees does not matter that much.

The theoretical finance literature also contains a large number of models, usually with a single risky asset, where steep demand curves are exogenously assumed.[8] While this is amply justified by empirical evidence, it ignores the contradiction with neoclassical multi-asset benchmarks such as the CAPM and APT. In contrast, the sole purpose of our model is to produce such steep demand curves as an endogenous equilibrium outcome.

Multi-asset equilibrium models such as Admati (1985) and Merton (1987) face the same problem as the CAPM. Whenever the cross-sectional pricing of stocks is determined by the same investors who collectively hold the entire market portfolio, *clearly uninformed* supply shocks can no longer move alphas by more than a negligible amount, so demand curves will have to be horizontal.[9]

Our model may resemble a multi-asset generalization of information cost or participation cost models including Grossman and Stiglitz (1980), Grossman and Miller (1988), and Allen and Gale (1994). However, these models do not allow an individual to start managing money for anyone else, even after he has paid a fixed cost to become informed; in contrast, our model explicitly allows this, because it seems more plausible that professional investors are not limited to managing their personal wealth. This reveals the crucial impact that delegation has on asset pricing, while also allowing us to calibrate the model to an observable quantity (percentage fee). Our model also shares resemblance to Berk and Green (2004) where in equilibrium active funds have to earn their fees, but their paper focuses on the

[8] This covers virtually all single-asset models where agents are not risk-neutral and thus their risk aversion plays a role in pricing (e.g. Chen, Hong, and Stein (2002), Allen and Gale (1994), and many others). Also some multi-asset models (e.g. Barberis and Shleifer (2002) and Wurgler and Zhuravskaya (2002)) exogenously assume steep demand curves.

[9] Hence, exogenous tastes for individual stocks as in e.g. Fama and French (2004) can only produce negligible deviations from CAPM pricing. Gomes, Kogan, and Zhang (2003) offer an example of a multi-asset equilibrium where interesting price effects emerge from a conditional CAPM but where demand curves are still horizontal. In Daniel, Hirshleifer, and Subrahmanyam (2001), systematic risk can be mispriced but again individual stocks cannot meaningfully deviate from factor pricing.

dynamics of the mutual fund industry while our paper concentrates on equilibrium prices of stocks in the presence of active and passive funds.[10]

Our paper makes two main contributions. First, it presents the first generally applicable explanation for downward-sloping demand curves which gets the magnitude of the effect approximately right. Thus it provides a theoretical justification for the models that have exogenously assumed steep demand curves. Second, it illustrates that financial institutions do indeed matter for asset pricing. This is in contrast to all models based on a single representative agent, suggesting that such models may be better suited for pricing systematic risk than a wide cross-section of stocks with idiosyncratic risk. Furthermore, we obtain our result entirely without agency issues, complementing the existing literature (e.g. Ross (1989) and Allen (2001)) which has pointed out the relevance of institutions to asset pricing due to agency issues.

The paper proceeds as follows. Section 2 starts with a simple CAPM benchmark and contrasts it with empirical evidence to illustrate the puzzle. It also briefly addresses alternative hypotheses in the literature. Section 3 presents our model and the equilibrium, and it provides a numerical calibration to show the magnitudes of the predicted effects. Section 4 presents the other empirical predictions of the model. Section 5 discusses interpretations and extensions of the model, and section 6 concludes. All algebra is in a separate appendix available from the author.

2 The Puzzle: Theory and Empirical Evidence

No equilibrium model of course literally implies that the demand curve for a stock is perfectly horizontal.[11] The real question here is about the magnitude of the 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.

---

[10] The relationship to Berk and Green (2004) is further discussed in section 5.2.

[11] When the representative investor buys more of a stock, that stock becomes a larger part of his systematic portfolio risk, i.e. its beta increases, and thus it requires a higher return. However, in a well-diversified portfolio, the stock should represent only a tiny fraction of the portfolio anyway, so this effect should be negligible.

2.1 A Simple CAPM Calibration for Demand Curves

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 stock i pays a liquidating dividend of $\widetilde{x}_i = a_i + b_i\widetilde{y} + \widetilde{e}_i$. Systematic shocks to the economy are represented by the unexpected return on the market portfolio $\widetilde{y} \sim N\left(0, \sigma_m^2\right)$, idiosyncratic shocks to the stock are denoted by $\widetilde{e}_i \sim N\left(0, \sigma_{e_i}^2\right)$, and a_i and b_i are stock-specific constants.[12] The return on the risk-free asset is normalized to zero.

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

The representative investor's maximization problem is:

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

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

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

The price is equal to the expected payoff a_i minus a discount, where 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 $\sigma_{e_i}^2 = 900$ for all stocks and $\sigma_m^2 = 0.04$ for the market variance. We start by letting the representative investor hold the entire market portfolio, so that $u_i = 1$ for all stocks. We also set $\gamma = 1.25 \times 10^{-5}$ 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%.

---

[12] Since the market return is a value-weighted return on individual stocks, the idiosyncratic stock returns have to add up to zero. We ignore this constraint for analytical convenience. This has a negligible impact on our results when there is a large number of assets.

Now consider a supply shock of -10% to a stock. Suppose, for example, that 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 only a tiny 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 even less: 0.11 bp. This is what the "almost perfectly horizontal" demand curves mean.[13]

What is the intuition for the result? In equilibrium, the representative investor is willing to bear a large amount of systematic market risk for a risk premium of 5%. Given that he holds large number of stocks (1,000), a 10% supply shock to an individual stock is only a tiny fraction of his entire portfolio (1/10,000). If he requires a 5% risk premium for an investment equal to the size of his entire portfolio, he will require only a tiny fraction of that premium for an investment equal to a tiny fraction of his entire portfolio.

2.2 Empirical Evidence for Demand Curves

To estimate the slope of the demand curve for a stock, most studies focus on large supply shocks where the source can be identified as uninformed both by 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.).[14] Trading due to merger arbitrage strategies also seems to produce a significant price impact (Mitchell, Pulvino, and Stafford (2004)) and could be used to extract elasticity estimates. Nevertheless, it is generally not easy to control for the information conveyed

[13] These results are not affected by the choice of CARA utility as opposed to CRRA utility. We further document this in a separate appendix (available from the author).

[14] In fact Scholes does find a significant price effect following block trades, but it seems almost unrelated to the size of a transaction. Since the cross-sectional dispersion in the price effect is large and related to the identity of the trader, a relationship between trade size and trader identity might account for his finding. His paper does not show results within subgroups for different types of investors.

by these events, and this could contribute to the relatively wide dispersion in elasticity estimates across different papers.[15]

A cleaner approach involves changes in widely tracked stock market indexes. Shleifer (1986) uses changes in the S&P 500 index and the consequent demand shocks by investors tracking the index to measure the slope of the demand curve. Several other papers have followed this approach and documented a substantial price impact around S&P 500 index changes (e.g. Lynch and Mendenhall (1997)) which seems to have grown with the popularity of indexing (Morck and Yang (2001)). Similar effects have been documented for other indexes in the U.S., such as the Russell indexes, as well as for a variety of indexes around the world. The studies for the S&P 500 suggest a price elasticity of demand of approximately unity. In recent years there has been an approximately 10% cumulative price impact for index additions and deletions with an annual peak of 15% in 2000 (Petajisto (2004)), while the demand shock by mechanical indexers has been approximately 10% of the shares outstanding of each stock.[16]

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 0.001% price impact for a 10% demand shock, and adjusting 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.

2.3 Alternative Hypotheses for the Evidence

It should be emphasized that currently there is no general explanation for steep demand curves. However, in the specific context of index additions and deletions, there are several hypotheses to explain the evidence. Yet none of the papers in the literature has attempted to calibrate these hypotheses to actual data. Could they theoretically explain a significant fraction of the index premium? How applicable are they across all the index evidence?

---

[15] A particularly amusing example of downward-sloping demand curves is provided by Rashes (2001) who finds significant price impacts even for trades where investors appeared to be confused about ticker symbols and traded a wrong stock.

[16] 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)).

2.3.1 Liquidity

Stocks in the S&P 500 are typically among the most liquid stocks, which shows in their greater trading volume and narrower bid-ask spreads. Perhaps liquidity creates a price premium for these stocks, along the lines of Amihud and Mendelson (1986). If S&P 500 membership per se increases liquidity, this would explain at least some price impact around index changes.

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.[17] Greenwood (2004) studies a large event for the Nikkei 225 index which had a significant price impact on the stocks that were in the index before and after the event. When MSCI redefined its indexes (tracked closely by $600 billion and loosely by $3 trillion) to be based on the float and not the number of shares outstanding, many practitioners were taking speculative positions in anticipation of intra-index price effects.[18] Liquidity, as arising from index membership per se, cannot account for all these findings.

2.3.2 Market Segmentation

Merton (1987) suggests that the price of a stock is increasing in its investor base. Applying his reasoning to our context, the addition of a stock to the S&P 500 could increase its visibility to investors and make information more widely available. This could then push up the stock price.

While this explanation could contribute to the effect, it faces the same challenge as the liquidity hypothesis with the intra-index events. It is easy to believe that the investor recognition of a stock depends on membership in the S&P 500 or on market capitalization, but it is much harder to explain why the official index weight would matter for investor recognition once market capitalization and index membership have already been taken into account.

---

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

[18] "MSCI's Stock Shuffle Turns Managers Into Stock Pickers," *Wall Street Journal*, 11/30/2001.

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 of section 2.1 with 20 stocks instead of 1,000, we still get only a 0.05% price impact. Even this exposure to market risk is so large that it implies a very low risk aversion for investors and almost perfectly horizontal demand curves.

Van Nieuwerburgh and Veldkamp (2006) present a model with an extreme form of market segmentation where each investor learns about and trades only one stock. This would in fact be sufficient to generate steep demand curves for individual stocks. However, it comes at the high cost of requiring that none of the informed investors is active in more than one stock – something that is difficult to reconcile with the actual practice of institutional money managers. The authors discuss ways to relax this, but anything that generates less extreme forms of market segmentation simultaneously reintroduces the puzzle about steep demand curves.

2.3.3 Information

Addition to the S&P 500 may convey positive information about a stock, as suggested by e.g. Denis et al. (2003). But for information to be the sole explanation, we again run into the challenge of the intra-index price effects. Other evidence can be obtained from indexes such as the Russell 2000 where membership is based on a mechanical market-cap rule, and yet we still observe both economically and statistically significant price effects (e.g. Petajisto (2004)). Practitioners also keep a close eye on changes to other mechanically determined indexes such as the Nasdaq 100.[19] In fact even for the S&P 500, the bureaucratic index changes in July 2002 represent a clearly uninformative event which nevertheless produced the usual magnitude for the price impact.[20]

---

[19] "Nasdaq 100 Index Shuffle Is Expected to Bring 13 Changes to List of Stocks," *Wall Street Journal*, 11/12/2001.

[20] Barberis, Shleifer, and Wurgler (2004) point out that index membership may in fact change the beta of a stock. This could potentially lead to a price impact around index changes. However, the beta of a stock cannot change due to index membership unless mechanical fund flows are able to influence prices, i.e. unless demand curves slope down. Hence, any such change in beta should be taken as evidence of downward-sloping demand curves, but of course it leaves open the question about why demand curves slope down in the first place.

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 a 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. 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.[21]

Such institutions have emerged presumably because there is some fixed cost to becoming an informed and active market participant. Uninformed "end investors" then 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.[22] For 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.

3.2 The Model

We consider a setting (Figure 2) similar to the CAPM calibration in section 2.1. The main difference is an explicit layer of institutions between end investors and the stock market: the end investors can invest in the stock market only indirectly through an active manager

[21] For smaller and transitory order imbalances, it would be realistic to consider the impact of market makers on the slopes of demand curves. However, membership changes in the S&P 500 represent very large and permanent supply shocks (and their price impacts persist even after several months), so they have to be primarily accommodated by other investors with longer investment horizons. Since we are interested in price effects that last for months or years, we ignore market makers altogether.

[22] This is perhaps the most commonly quoted value for the annual fee, but there is some dispersion here. For example, Kacperczyk, Sialm, and Zheng (2004) 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.

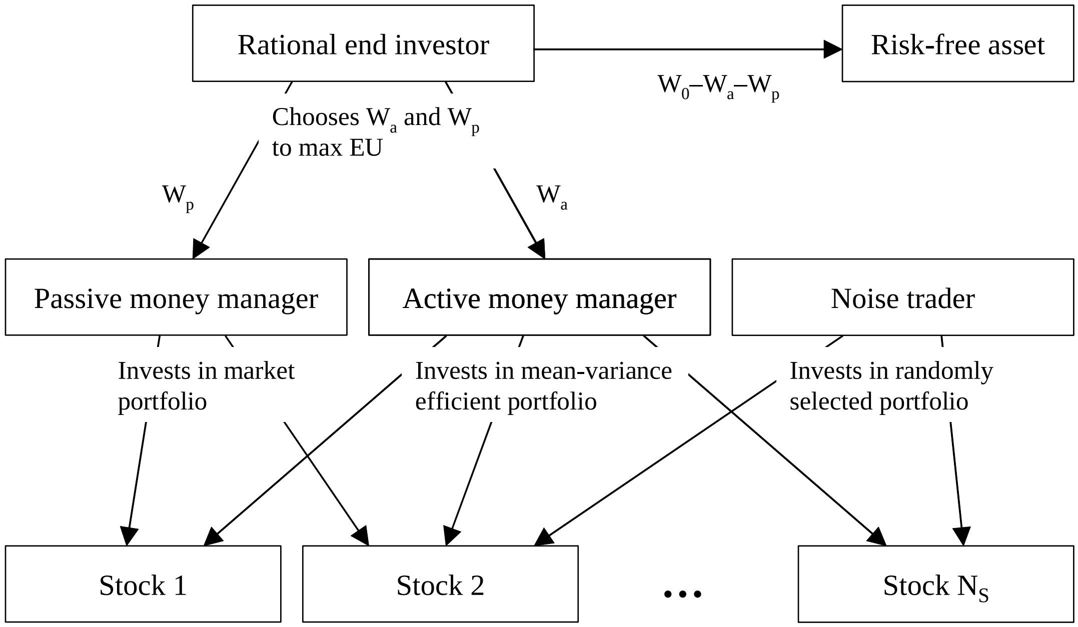

Figure 2: The basic setup for the model.

(a stock picker) and a passive manager (who just holds the market portfolio). We also assume there are exogenous noise traders who hold a randomly chosen portfolio of stocks.[23] Since the noise traders deviate from the market portfolio, they create profitable trading opportunities for the active managers. 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 stock i pays a liquidating dividend of $\widetilde{x}_i = a_i + b_i\widetilde{y} + \widetilde{e}_i$ dollars. Systematic shocks to the economy are represented by the unexpected return on the market portfolio $\widetilde{y} \sim N\left(0, \sigma_m^2\right)$. Idiosyncratic shocks to the stock are denoted by $\widetilde{e}_i \sim N\left(0, \sigma_{e_i}^2\right)$. a_i and b_i are stock-specific constants. The return on the risk-free asset is normalized to zero.

To keep the mathematics simple while allowing for a large number of stocks, we make two assumptions. We let all stocks have the same values of a_i, b_i, and $\sigma_{e_i}^2$. We also assume

[23] Alternatively, we could assume an unobservable noisy supply for each stock. Since we will be calibrating the model to plausible parameter values and since we will have investors who hold the market portfolio, it is more convenient to talk explicitly in terms of noise trader holdings.

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 γ_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:

$$\max_{\{W_a, W_p\}} E\left[-\exp\left(-\gamma_e \widetilde{W}_1\right)\right]$$
$$\text{s.t.} \quad \widetilde{W}_1 = W_0 + W_a \widetilde{R}_a + W_p \widetilde{R}_m, \tag{3}$$

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

To write out the return on the active portfolio, we need to know the "cost" of the portfolio, i.e. how much capital it ties up. In reality all risky positions tie up a positive amount of capital – even short-only funds are constrained in their positions by the amount of capital they have. To capture this notion, we do not allow short positions (or borrowing) to finance long positions. We assume the cost of the active portfolio is given by its long positions only:

$$\sum_{v_i > 0} v_i = 1. \tag{4}$$

This represents a collateral requirement where all the cash generated by short sales is invested in the risk-free asset, which is also a reasonable approximation to reality.[24,25]

Denoting the excess return on stock i as $\widetilde{R}_i$ and the price of the market portfolio as P_m,

[24] Investors are usually required to deposit 102% of the cash proceeds of the short sale with their broker (D'Avolio (2002)).

[25] The cost of the active portfolio in equation (4), including the size of the collateral constraint and the extent of leverage, can actually be selected from a wide class of allowable cost functions. The exact choice of our definition matters only for the value of the percentage fee f: if the end investor needs to commit only a small amount of capital to establish his active positions, the same fixed cost C will produce a higher percentage fee f, and vice versa if the end investor needs to commit a large amount of capital.

we can then write the portfolio returns as

$$\widetilde{R}_a = \left(\sum_{i=1}^{N_S} v_i \widetilde{R}_i\right) - f \tag{5}$$

$$\widetilde{R}_m = \frac{1}{P_m}\sum_{i=1}^{N_S} P_i \widetilde{R}_i, \tag{6}$$

so the active portfolio has weights v_i and a constant proportional fee f 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 $\widetilde{R}_a = \alpha_a + \beta_a \widetilde{R}_m + \widetilde{\varepsilon}_a$ where β_a is the market beta of the portfolio and $\widetilde{\varepsilon}_a \sim N\left(0, \sigma_a^2\right)$. Then the after-fee abnormal return α_a and the idiosyncratic variance σ_a^2 of the manager's portfolio are given by:

$$\alpha_a = \sum_{i=1}^{N_S} v_i \alpha_i - f \tag{7}$$

$$\sigma_a^2 = \sum_{i=1}^{N_S} v_i^2 \sigma_i^2 \tag{8}$$

where α_i and σ_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.

3.2.3 Active Managers

An active manager offers the end investor a portfolio with stock weights v_i (some of which may be negative) and a proportional fee f. We assume that there is a market for active managers: anyone can become an active manager by paying a fixed dollar cost C. It allows the manager to learn the stock-specific parameters a_i, b_i, and $\sigma_{e_i}^2$ and then actively pick an efficient portfolio. The manager recovers this fixed cost by imposing a fee which is a constant percentage of assets under management.[26]

[26] Note that it would be very difficult to maintain any other kind of fee structure in equilibrium. Since portfolios are virtually costless to repackage, any nonlinear pricing (including nonlinear fees) would represent an arbitrage opportunity. Not surprisingly, linear fee structures also appear to be the norm in practice.

Active managers compete with one another to provide the end investor with a portfolio that maximizes his expected utility (3), subject to the constraint that the managers have to earn their costs at the end investor's optimal allocation $W_a = W_a^*$.

Given the highly simplified structure of this model, it is worth commenting on two of its features. First, since we assume a fixed dollar cost but no offsetting diseconomies of scale, in equilibrium with free entry there will be only one active manager whose total fee is exactly enough to cover his fixed cost C. If the manager's fee exceeds his cost, someone else will step in, undercut the fee of the incumbent, and win the business of all end investors.[27] In reality we of course observe a large number of competing yet coexisting actively managed funds, even within relatively narrow market segments, which suggests the presence of some diseconomies of scale.[28] While it would be realistic to include these considerations in the model, our main objective is to find out how the intermediaries and their proportional fee affect the cross-sectional pricing of assets, and here a simpler structure for the money management industry should keep our main result as transparent as possible.

Second, active managers in reality tend to combine their active positions in individual stocks with a large position in the market portfolio. In fact, many of them think of their portfolios as consisting of 100% investment in the benchmark index plus a long-short overlay portfolio which contains their active positions. Later in section 5.3 we explicitly consider a setting where active managers also take a large position in the market portfolio, and this makes our results much stronger. However, in the interest of simplicity, the main presentation of our model here assumes active managers do not serve such a dual role; instead they only take active positions in individual stocks.

The dollar amount an active manager invests in stock i is $W_a v_i$, so he offers portfolio weights v_i (e.g. $v_i = 10\%$ in stock i) which are then scaled by the dollar amount W_a invested

---

We abstract away from return-based incentive fees since 98% of U.S. mutual funds do not have such fees (Elton, Gruber, and Blake (2003)).

[27] Perhaps more realistically, we could divide the economy into n segments (industries), each with a fixed cost of $\frac{C}{n}$. In equilibrium we can then have n active managers who each specialize in one segment.

[28] Chen, Hong, Huang, and Kubik (2004) discuss the organizational diseconomies of an actively managed fund. They find empirical evidence that such diseconomies do erode fund performance. Alternative approaches are presented by Hortacsu and Syverson (2004) who suggest search costs to explain the existence of a large number of funds (including funds with different fees yet virtually identical portfolios), while Mamaysky and Spiegel (2002) suggest that multiple funds could exist to cater to investors' heterogeneous preferences.

by the end investor. In other words, the dollar size of the manager's active position in each stock is directly proportional to his total amount of assets under management.

3.2.4 Equilibrium between End Investor and Active Manager

The end investor chooses his optimal allocations W_a and W_p, taking the excess returns on the market portfolio and the active portfolio as exogenous. We can write this problem as:

$$\max_{W_a, W_p} E\left[u\left(W_0 + W_p\widetilde{R}_m + W_a\left(\sum_i v_i\widetilde{R}_i - f\right)\right)\right]. \tag{9}$$

Since the optimal allocations will depend on the manager's choices $\{v_i\}$ and f, we can write them as functions $W_a^*(\{v_i\}, f)$ and $W_p^*(\{v_i\}, f)$.

After some algebra, we obtain the optimal allocations to the active and passive portfolios:

$$W_a^* = \frac{E\left[\widetilde{R}_a\right] - \beta_a\eta}{\gamma_e\sigma_a^2} = \frac{\alpha_a}{\gamma_e\sigma_a^2} \tag{10}$$

$$W_p^* = \frac{E\left[\widetilde{R}_m\right]}{\gamma_e\sigma_m^2} - \beta_a W_a^* = \frac{\eta}{\gamma_e\sigma_m^2} - \beta_a W_a^*, \tag{11}$$

where η denotes the market risk premium. When we plug these expressions into the end investor's maximization problem (9), we can write the objective function in terms of the certainty equivalent of the end investor:

$$W_0 + \frac{1}{2\gamma_e}\left[\underbrace{\left(\frac{\eta}{\sigma_m}\right)^2}_{\substack{\text{Sharpe ratio} \\ \text{of market}}} + \underbrace{\left(\frac{\alpha_a}{\sigma_a}\right)^2}_{\substack{\text{appraisal ratio of} \\ \text{active portfolio}}}\right]. \tag{12}$$

The end investor's expected utility thus depends on the Sharpe ratio of the market and the appraisal ratio of the active portfolio. This is consistent with Treynor and Black (1973) and the subsequent investment literature which advocate the appraisal ratio (also known as the information ratio) as an appropriate objective for an active manager.

The active manager then chooses portfolio weights $\{v_i\}$ and the fee f to maximize his appraisal ratio subject to the constraint that he cover his fixed cost. Consequently, his portfolio weights will be linear in alpha:

$$v_i = \left(\frac{1}{\sum_{\alpha_j>0}\frac{\alpha_j}{\sigma_j^2}}\right)\frac{\alpha_i}{\sigma_i^2}. \tag{13}$$

These are also the same portfolio weights the end investor (or any other mean-variance investor) would choose himself if he was trading stocks directly.[29] Given these weights, the fee f is simply the lowest percentage that will still allow the manager to cover his fixed cost.

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

$$W_i = W_a v_i = \frac{W_a}{\sum_{\alpha_j>0} \frac{\alpha_j}{\sigma_j^2}} \frac{\alpha_i}{\sigma_i^2} = \frac{\alpha_i}{\gamma \sigma_i^2}, \tag{14}$$

where we defined the "effective risk aversion" of the active manager as

$$\gamma = \frac{1}{W_a} \sum_{\alpha_j>0} \frac{\alpha_j}{\sigma_j^2}. \tag{15}$$

This is the implied coefficient of absolute risk aversion of the active manager if he was a CARA investor investing his own wealth.[30] Since 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. Yet this notation is very useful, as it simplifies our equations and offers a convenient interpretation in the equilibrium analysis.

3.2.5 Market Clearing

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

$$u_p + u_{in} + u_i = 1. \tag{16}$$

In equilibrium the active manager has $u_i P_i$ dollars in stock i. Equating this with his dollar demand from equation (14), we find the equilibrium alpha as a linear function of the

[29] We present another more formal derivation (together with the derivation of all other formulas in this paper) in a separate appendix available from the author.

[30] The manager's true personal risk aversion is not even defined, as he has no personal wealth or utility function.

market-clearing supply u_i:

$$\alpha_i = \frac{\gamma \sigma_{e_i}^2}{P_i} u_i. \tag{17}$$

Hence, the manager will be long positive-alpha stocks and short negative-alpha stocks. The equilibrium price of stock i will then be

$$P_i = \underbrace{a_i}_{\substack{\text{expected} \\ \text{payoff}}} - \underbrace{b_i \eta}_{\substack{\text{discount for} \\ \text{market risk}}} - \underbrace{\gamma \sigma_{e_i}^2 u_i}_{\substack{\text{deviation} \\ \text{from CAPM}}} . \tag{18}$$

By construction, the market portfolio will always have an alpha of zero. This implies that $u_i \sim N\left(0, \sigma_u^2\right)$. 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 η, as well as the fee f and the effective risk aversion γ 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:*

$$\eta = \frac{\gamma_e \sigma_M^2}{N_S a - \gamma_e \sigma_M^2} \tag{19}$$

$$W_p = N_S a - \gamma_e \sigma_M^2 \tag{20}$$

$$W_a = \frac{N_S \sigma_u}{2} \left[\sqrt{\frac{2}{\pi}} (a - b\eta) - \gamma \sigma_e^2 \sigma_u \right] \tag{21}$$

$$\gamma = \gamma_e + \frac{C}{N_S \sigma_e^2 \sigma_u^2} \tag{22}$$

$$f = \frac{2C}{N_S \sigma_u \left[\sqrt{\frac{2}{\pi}} (a - b\eta) - \gamma \sigma_e^2 \sigma_u \right]} \tag{23}$$

Here σ_M^2 denotes the dollar variance of the market portfolio. We keep the expressions simple by leaving some of them in terms of η or γ, both of which are endogenous variables.

3.3 Analysis of Equilibrium

3.3.1 Selection of Parameters

The model has essentially three free and meaningful parameters to pick: the length of the time period, the active manager's fixed cost C (which produces a fee f), and the dispersion in noise traders' holdings σ_u. For the rest of the parameters we either get reasonably good estimates from actual data (the market risk premium and volatilities) or they do not matter for our results (price normalization or the exact number of stocks). The model's restrictions then determine the joint equilibrium distributions of u_i, P_i, and α_i, which in turn determine the slope of the demand curve for a stock.

In the first calibration, we want to be as close as possible to the CAPM benchmark of section 2.1. 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 $\gamma_e = 1.25 \times 10^{-5}$ (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 $\beta_m = 1$), $\sigma_M^2 = 4 \times 10^8$ (to get a standard deviation of 20% for the market return), $\sigma_e^2 = 900$ (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). We again investigate the price impact of an exogenous -10% supply shock which would correspond to a stock being added to the S&P 500. We then perform the same calibration with the time period set to five years instead of one year.

3.3.2 Calibration Results

The expression for the effective risk aversion of the active manager as a function of the percentage fee f perhaps most clearly reveals the unique feature of our equilibrium:

$$\gamma = \gamma_e + \frac{C}{N_S \sigma_u^2 \sigma_e^2} \tag{24}$$

$$= \left(\frac{1}{1+\frac{f}{2}}\right)\left(\gamma_e + \frac{1}{\sqrt{2\pi}}\left(\frac{a-b\eta}{\sigma_e^2 \sigma_u}\right) f\right) \tag{25}$$

$$\approx \gamma_e + \frac{1}{\sqrt{2\pi}}\left(\frac{a-b\eta}{\sigma_e^2 \sigma_u}\right) f. \tag{26}$$

In other words, γ is approximately linear in the percentage fee f (and exactly linear in the fixed cost C). If the fee charged by the active manager is zero, then the active manager's risk aversion will match that of the representative end investor, and the model collapses to the CAPM benchmark. Exactly as before, a -10% supply shock to a typical stock will increase the price of the stock by only 0.11 basis points. However, the fee f has a very significant first-order effect on γ – even a tiny fee of 0.1% would increase γ by a factor of 40. Panel A in Table 1 illustrates the effect of the fee on the equilibrium distribution of

Table 1: The effect of the management fee.

Panel A: One-year horizon			
fee	95% confidence interval for α_i	effective risk aversion γ	price impact of a -10% supply shock
0	$[-0.0022\%, 0.0022\%]$	1.25×10^{-5}	0.0011%
0.1%	$[-0.08\%, 0.08\%]$	4.52×10^{-4}	0.04%
0.5%	$[-0.39\%, 0.39\%]$	2.21×10^{-3}	0.20%
1.0%	$[-0.77\%, 0.78\%]$	4.41×10^{-3}	0.40%
1.5%	$[-1.2\%, 1.2\%]$	6.61×10^{-3}	0.60%
2.0%	$[-1.5\%, 1.6\%]$	8.81×10^{-3}	0.80%
Panel B: Five-year horizon			
annual fee	95% CI: cumulative α_i over 5 years	effective risk aversion γ	price impact of a -10% supply shock
0	$[-0.011\%, 0.011\%]$	1.25×10^{-5}	0.0056%
0.1%	$[-0.40\%, 0.40\%]$	4.55×10^{-4}	0.20%
0.5%	$[-1.9\%, 1.9\%]$	2.20×10^{-3}	1.0%
1.0%	$[-3.7\%, 4.0\%]$	4.34×10^{-3}	2.0%
1.5%	$[-5.4\%, 6.0\%]$	6.42×10^{-3}	2.9%
2.0%	$[-6.9\%, 8.0\%]$	8.46×10^{-3}	3.8%

alphas, on the effective risk aversion, and on the price impact of a -10% supply shock with a one-year horizon. Panel B shows the same results with a five-year horizon.

With a one-year horizon and a realistic fee of 1.5% of assets under management, we get a price impact of 0.60%. This is orders of magnitude (over 500 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.

With a five-year horizon, the price effects are scaled up approximately by a factor of five. Now a -10% supply shock produces a price impact of about 3%, which is economically a very significant amount and roughly equal to one quarter of the actual S&P 500 index premium. While the crucial deviation from the CAPM arises solely due to the fee, the horizon also matters a great deal if we want to get close to the empirically observed price effect.

3.3.3 Intuition

Regardless of the horizon, our results are in stark contrast to traditional representative agent models where end investors' risk aversion shows up both in the pricing of market risk and in the pricing of idiosyncratic risk. In our setting, 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? The cross-sectional pricing of stocks is determined by the active manager who is constrained to invest exactly 100% of the wealth allocated to him by the end investors. In equilibrium the end investors will have to be indifferent between the actively managed portfolio, which has a positive alpha but charges a fee, and the passively managed portfolio, which has a zero alpha but also a zero fee. Hence, the before-fee alpha of the active portfolio has to be approximately equal to the fee. This in turn implies that the dispersion in the alphas of individual stocks has to be sufficiently wide in equilibrium to produce the nontrivial portfolio alpha. The dispersion in alphas thus represents an equilibrium level of "inefficiency" in the market, measured with respect to the active manager's information set.[31,32]

The alpha curve for a stock and the equilibrium distribution of alphas in the entire population of stocks are shown in Figure 3. The dotted lines indicate the typical positive and negative stock positions of the manager: he earns an alpha of 0.75% on each, adding

[31] Here the stock market is not "efficient" in the traditional sense because an active manager can pick stocks that outperform the market. But since this outperformance cannot be obtained without a cost and in equilibrium the cost largely eliminates the gains from outperformance, we could reasonably define this market as efficient.

[32] Alternatively, we could write the stochastic discount factor of the economy as

$$\widetilde{m} = 1 - \frac{\eta}{\sigma_m^2}\widetilde{y} - \gamma \sum_{i=1}^{N_S} u_i \widetilde{e}_i.$$

The first random term accounts for the systematic discount of a stock due to market risk (the CAPM price). The remaining random terms account for the idiosyncratic mispricings of individual stocks. There is no structure to these mispricings – only the active managers conducting fundamental analysis of individual firms are able to identify them.

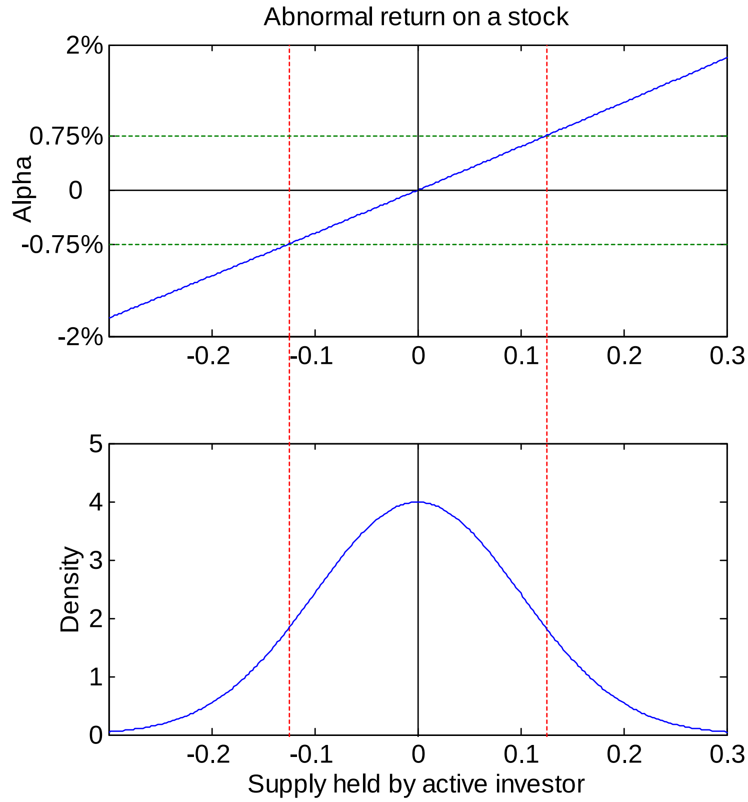

Figure 3: The alpha of an individual stock, and the distributions of alphas and active manager's stock holdings (u_i) in the entire population of stocks.

up to a portfolio alpha of 1.5% which just covers the fee of 1.5%. The slope of the alpha curve is now about 500 times greater than in the CAPM benchmark.

Regardless of the horizon, the distribution of annual alphas is the same. Yet the pricing results are very different, because the alphas across the entire period are capitalized into prices today (Figure 4). With a one-year horizon, a 1% annual alpha translates to a 1% underpricing, but with a five-year horizon the same 1% annual alpha translates to a 5% underpricing.

3.3.4 Interpretation of Horizon

How should we interpret the horizon of the model? In a one-period model, the horizon is essentially a period of time after which prices fully converge to their fundamental values. Yet in reality, no such convergence is guaranteed for stocks. It is thus better to think

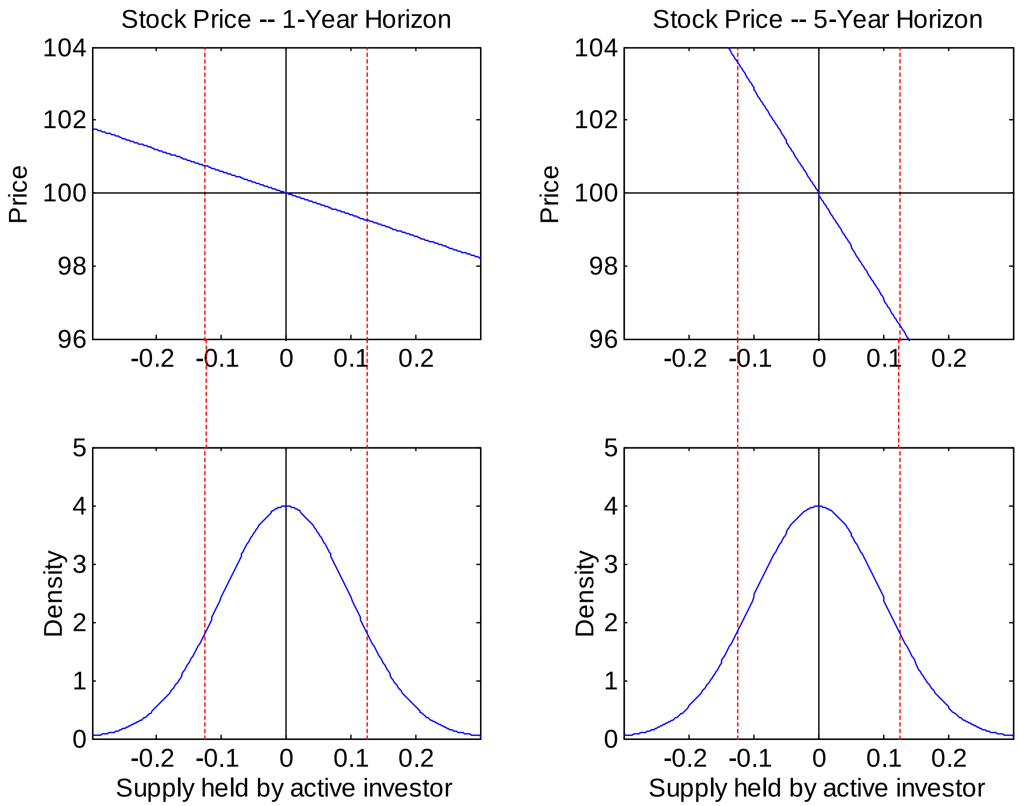

Figure 4: The price of an individual stock, and the distributions of prices and active manager's stock holdings in the entire population of stocks, for 1-year and 5-year horizons. The distribution of annual alphas is the same in each case, but since the alphas over the entire period are capitalized into prices today, the longer horizon scales up the "mispricings" today.

in terms of the expected half-life of a mispricing – e.g., the five-year horizon should be interpreted as an expected half-life of 2.5 years for a mispricing.

The choice of an appropriate horizon then becomes primarily an empirical question. DeBondt and Thaler (1985) find slow mean reversion in returns over a three-to-five-year period, while Jegadeesh and Titman (2001) find momentum at a one-year horizon and a partial or full reversal (depending on the sample period) over the following four years. This suggests that mispricings may indeed take several years to reverse. Cohen, Gompers, and Vuolteenaho (2002) construct a VAR model which allows them to estimate the reversal of a pure expected-return shock. Their results indicate a half-life of at least 2.5 years (figures 2 and 7 in their paper).

In the context of index changes, the price impact seems to last at least for two months, but beyond that our tests start to lose power to distinguish between alternatives. Professional investors seem to have divergent views on this topic, with some of them believing a stock will have a permanent premium as long as it stays in the index. Certainly a full one-year reversal seems implausible, as it would offer easy opportunities to earn 10% annual alphas. More generally, if a moderate mispricing can exist today, how can we be so sure it cannot exist tomorrow?

Overall, a half-life of 2.5 years for a mispricing seems roughly consistent with empirical evidence, so we adopt the five-year horizon as a reasonable compromise. The main virtue of the one-year horizon is that it makes the numbers in the calibration a little more transparent.

3.3.5 The Model and Reality

The model's predicted 3% price impact for S&P 500 index changes is in fact unrealistically low for the effect we describe. We have assumed frictionless short selling, and consequently the actively managed portfolio turned out to be a market-neutral long-short portfolio. In reality mutual funds and many other institutional investors almost never take short positions and they carry significant exposure to systematic market risk.

Section 5.3 presents a more realistic model where the active manager combines his active long-short portfolio with a passive investment in the market portfolio. This makes demand curves even steeper, and with plausible parameter values the price impact of S&P 500 index addition increases from 3% to 14% (see Table 2 on page 33). Here the same percentage fee now represents a greater fraction of the active positions because now the portfolio also includes a large passive position.

While we should typically not expect a simple model to be an accurate predictor of real-life price impact, the numbers from our calibrations should be taken as evidence that the mechanism we describe is economically significant and has the potential to explain a large part of the empirically observed price effects.

4 Empirical Implications

4.1 Predictions

The most immediate testable prediction of the model is the overall magnitude of the slopes of demand curves under reasonable parameter values. This was already discussed in the numerical calibration of the previous section.

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

$$P_i = a_i - b_i\eta - \gamma\sigma_{e_i}^2 u_i \tag{27}$$

$$\gamma \approx \gamma_e + \frac{1}{\sqrt{2\pi}}\left(\frac{a - b\eta}{\sigma_e^2\sigma_u}\right) f. \tag{28}$$

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.[33] 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 (γ) of the active manager. The price elasticity of demand for stock i is then

$$\frac{\frac{dQ_i}{Q_i}}{\frac{dP_i}{P_i}} = \frac{\frac{du_i}{1}}{\frac{dP_i}{P_i}} = P_i\frac{du_i}{dP_i} = -\frac{P_i}{\gamma\sigma_{e_i}^2}. \tag{29}$$

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,[34] 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.*

---

[33] Note that the deviation is sometimes positive and sometimes negative (depending on the sign of u_i), so idiosyncratic risk alone will not be linked to expected returns.

[34] In fact, if there is no segmentation, then small firms (measured by operating size such as revenues) will always command a smaller risk premium in equilibrium, giving rise to an inverse size effect. When the market is segmented, it is possible to maintain a relatively constant density of investors in each stock. We address these issues explicitly in a separate appendix to this paper.

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 has to be the same as the active managers still earn their fees, but now the same dispersion of alphas exists over a wider range of the managers' stock holdings, so the change in alpha (and price) for a supply shock of a given size is smaller. In other words, a noise trader can minimize his own price impact by trading in stocks where the volatility of aggregate noise trader holdings is high.

Our model also implies that noise traders can move prices, and in fact they can increase the volatility of a stock beyond the volatility of its fundamentals.

Implication 5 *Stocks with a greater volatility of noise trader holdings will exhibit greater price volatility, unless the shocks to noise trader holdings are inversely correlated with fundamental news.*

4.2 Evidence

The link between active management fees and the slopes of demand curves is tested in a separate paper (Petajisto (2004)), which provides empirical evidence from the large-cap and small-cap segments of the market using data from S&P 500 and Russell 2000 index changes. It finds that small-cap stocks exhibit steeper demand curves than large-cap stocks, which is consistent with the higher management fees of active small-cap mutual funds. Naturally, it would be interesting to test this prediction even more broadly across various market segments or multiple countries.

The predicted cross-sectional link between idiosyncratic risk and demand curves is strongly confirmed by empirical tests for both indexes (Petajisto (2004)).

5 Interpretations and Further Discussion

5.1 Grossman and Stiglitz (1980) and the Necessity of Institutions

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).[35] Could we perhaps use their model, or its multi-asset extensions such as Admati (1985) or Biais, Bossaerts, and Spatt (2006), 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 absence of new information. For example, when a stock is added to the S&P 500, every active trader in the stock who is not consciously ignoring news will know who the new buyers are and why the stock price went up. Thus any price effect from index addition would have to come from the risk aversion of the investors and not the rational expectations story of the model.

Let us then investigate a modified multi-asset version of the Grossman-Stiglitz model to see if it would fit better. Assume that in a large cross-section of stocks, the uninformed investors are completely passive and thus have a perfectly inelastic demand.[36] Prices are then exclusively set by the informed investors.

---

[35] Also Allen and Gale (1994) come remarkably close to Grossman and Stiglitz (1980). Similarly, the most fundamental difference between our paper and theirs (namely, the delegation of portfolio management) is the same.

[36] When the cross-section of firms exhibits wide dispersion in operating sizes and scaled-price ratios, these simple measures become virtually useless for the time-series trading of an individual stock. Without more

To generate the same slope for the demand curve as in our model with a fee of 1.5%, the informed investors would have to have a collective risk aversion equal to the effective risk aversion of our active manager (Table 1 on page 21). Since this is over 500 times the absolute risk aversion of all investors in the economy, it implies that one investor out of 500 would choose to become informed. Essentially this investor faces a trade-off: either he is uninformed and holds a tiny fraction of the market portfolio, or he becomes informed and suddenly takes large enough positions to accommodate all the demand shocks due to noise traders.

It seems like a stretch to say that this huge increase in his risky portfolio (about 40-fold in our calibration) comes 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. Certainly the incentive of an informed investor to sell money-management services to others is considerable.

This takes us to the central issue: 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 other investors are willing to allocate to him.[37] Yet the Grossman-Stiglitz setting effectively assumes even the informed investors still keep investing their own wealth but they just borrow massively to finance their very large portfolios. Thus the model is missing the crucial part of the mechanism which is the trade-off of end investors (uninformed investors) when allocating wealth to active managers (informed investors) and the resulting equilibrium value for the effective risk aversion of the active managers.

Other rational expectations models such as Admati (1985) and Biais, Bossaerts, and Spatt (2006) face similar difficulties. First, they cannot explain why the *clearly uninformed* supply shocks would have a price impact, because even the uninformed rational investors

---

detailed stock-specific information, the uninformed investors can therefore only have an almost perfectly inelastic demand for an individual stock.

Biais, Bossaerts, and Spatt (2006) actually do construct a price-contingent strategy for uninformed investors, but only at the level of six Fama-French portfolios and not at the level of individual stocks, thus sidestepping the issue mentioned here.

[37] Note that the manager's personal risk aversion is unaffected, but in order to take positions on behalf of his investors, he would have to increase his risky positions.

could trade against them. Second, assuming that *only* the informed investors trade against *any* supply shocks, we could generate steep demand curves by simply assuming the informed investors are extremely risk averse, but then again we have a hard time explaining why such investors would not start managing money for *other* investors, given the significant Sharpe ratios they can generate.

Hence, to answer our question about equilibrium slopes of demand curves, we do indeed need something like our model where the delegation of portfolio management is made explicit. Costly information acquisition, conducted by individual investors directly, would be very hard to reconcile with a plausible multi-asset equilibrium.

5.2 Applying Berk and Green (2004) for Asset Pricing

Our story also shares many features with Berk and Green (2004), which is a model of asset flows and dynamics in the mutual fund industry. Both models assume that active managers have some skill to begin with. On a *net* return basis, i.e. after fees and trading costs, Berk and Green assume that investors are indifferent between active funds and passive index funds, whereas we derive the same condition in equilibrium. They similarly calibrate their active funds to an annual fee of 1.5%. Their model is obviously not intended for asset pricing, as all returns are exogenously given, but perhaps we could slightly modify it to explain equilibrium asset prices?

Berk and Green derive their results in large part from the quadratic dollar cost, which we can interpret as a linear price impact, faced by an active manager. Aware of his own price impact, the active manager determines the optimal size of his active positions and passively indexes *all* additional assets.

However, there are thousands of mutual funds and other institutional investors in the US equity market. Equilibrium pricing is determined by the *collective* price impact of all these funds, which is substantially larger than the price impact of an individual fund. Therefore we can greatly simplify the model by assuming price-taking behavior, which is the approach we choose in our model.

It also seems rather clear that giving more money to active managers will lead them to take larger active positions (in dollar terms),[38] thus pushing prices closer to their fundamental values. In fact this is the fundamental mechanism that supports equilibrium in

[38] See e.g. Cremers and Petajisto (2007) for direct empirical evidence on this.

our model – if the active managers get too much money, pricing becomes so efficient that their net alphas become negative, prompting investors to withdraw their money. Because Berk and Green assume a manager's active positions are completely unrelated to his assets under management, this assumption would have to be relaxed before using their model for equilibrium pricing.

Hence, the Berk-Green model would have to be modified for equilibrium pricing, taking into account these important conceptual issues. Our model, while constructed independently, serves that purpose.

5.3 Active Managers Benchmarked Against Market

In reality, the true fees on actively managed portfolios can be much higher than in our simplified model. Managers are benchmarked against a market index, so their active positions only consist of their deviations from the benchmark. If the active positions are smaller than the investment in the benchmark index, our model understates the effect of fees on the slopes of demand curves.

The high cost of active management has been documented by Miller (2007) and Cremers and Petajisto (2007). The former paper estimates active positions at only 15% of the total portfolio for large-cap mutual funds; the latter paper calculates aggregate active positions at about 30% of the total portfolio.[39] In either case, the magnification effect on fees is substantial. How should we adjust our model to take this into account?

Another manifestation of the same issue is that mutual funds almost never take short positions. Yet in our earlier calibration, the 95% confidence interval for the holdings of the active managers was $[-20\%, 20\%]$ of the supply of each stock. One natural way to eliminate almost all short positions is to let the managers also hold 20% of the market portfolio. This shifts their 95% confidence interval of holdings to $[0\%, 40\%]$, implying that they have aggregate short positions in only about 2.5% of stocks. This is a simple way to simultaneously address the issue of overstated active positions, so we adopt it for the subsequent analysis.

[39] Miller (2007) estimates the size of active positions from return data after assuming that the active long-short portfolio is as volatile as the benchmark index. Cremers and Petajisto (2007) compute the active positions directly from mutual fund holdings data.

More formally, we can change the manager's participation constraint to the following:

$$fW_a \geq C + (zN_S\sigma[v_i] - v_m)^2. \tag{30}$$

Here v_m is the portfolio weight on the market index, v_i is the active portfolio weight on stock i, $\sigma[v_i]$ is the cross-sectional standard deviation in the active portfolio weights, N_S is the number of stocks, and z is the z-value for the normal distribution (e.g., $z = 1.96$ if only 2.5% of stocks will be shorted in the aggregate by active managers). This additional cost implies that there exists an optimum mix of active portfolio weights with the market index weight. It could arise for a variety of reasons such as costly short sales or a specific tracking error objective for the manager.

The manager again starts by picking active weights that are linear in alpha:

$$v_i = k\frac{\alpha_i}{\sigma_i^2}. \tag{31}$$

To minimize his costs, the manager can then choose his market exposure as

$$v_m = zN_S\sigma[v_i] = kzN_S\sqrt{Var\left[\frac{\alpha_j}{\sigma_j^2}\right]}. \tag{32}$$

Since portfolio weights $v_m + \sum_{i=1}^{N_S} v_i$ must add up to 1, we obtain the normalization constant k:

$$k = \frac{1}{zN_S\sqrt{Var\left[\frac{\alpha_j}{\sigma_j^2}\right]} + \sum_{j=1}^{N_S}\frac{\alpha_j}{\sigma_j^2}}. \tag{33}$$

After some algebra, we can solve for the effective risk aversion of the active manager in equilibrium in terms of either the dollar cost C or the percentage fee f:

$$\gamma = \gamma_e + \frac{C}{N_S\sigma_e^2\sigma_u^2} \tag{34}$$

$$\approx \gamma_e + \frac{z(a - b\eta)}{\sigma_e^2\sigma_u}f \tag{35}$$

For a fixed dollar cost C, this version of the model produces the exact same demand curves for stocks as the basic model. But as a function of the percentage fee f, we get different results because the active manager no longer has a pure long-short active portfolio. Table 2 presents the calibration comparable to Panel B in Table 1. Comparing equations (26) and (35) for the manager's effective risk aversion, we see that the model with benchmarking scales up the price effect by a factor of $z\sqrt{2\pi} \approx 2.5z$.

Table 2: The effect of the management fee; five-year horizon and long-only portfolio.

annual fee	95% CI: cumulative α_i over 5 years	effective risk aversion γ	price impact of a -10% supply shock
0	$[-0.011\%, 0.011\%]$	1.25×10^{-5}	0.0056%
0.1%	$[-1.9\%, 2.0\%]$	2.18×10^{-3}	0.98%
0.5%	$[-8.6\%, 11\%]$	1.07×10^{-2}	4.8%
1.0%	$[-16\%, 23\%]$	2.09×10^{-2}	9.4%
1.5%	$[-21\%, 38\%]$	3.10×10^{-2}	14%
2.0%	$[-27\%, 56\%]$	4.09×10^{-2}	18%

For example, if the manager has an active portfolio with about \$20 in long positions and \$20 in short positions, now he also has \$100 invested in the benchmark index. Thus he is charging the percentage fee on \$100 worth of total assets and not only on \$20 (the active long positions) as before.[40] This effectively multiplies the slopes of demand curves and the price impact of index addition by a factor of five, from about 3% to 14%. This magnitude is close to and even slightly above the recent S&P 500 index premium of about 10%.

The index premium is of course only one manifestation of the broader issue of steep demand curves, so we should not focus exclusively on matching that. However, it is still reassuring that this effect is approximately within range of reasonable parameter values for the model, especially given that traditional pricing models are off by several orders of magnitude.

5.4 Transaction Costs

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

The first immediate 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 can even be less

[40] These numbers are also roughly in line with both Miller (2007) and Cremers and Petajisto (2007).

than 0.1%. Yet the S&P 500 premium has averaged about 10% which is certainly sufficient to produce abnormal returns even net of transaction costs. Moreover, some of the largest additions such as Goldman Sachs, UPS, and Microsoft have had the lowest transaction costs, yet they have experienced some of the largest price impacts.

The more fundamental challenge is that when end investors trade stocks directly, 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 empirical evidence on demand curves shows that prices (and alphas) change smoothly even beyond the transaction cost as we vary the size of the supply shock. A story based on transaction costs cannot match this key feature of demand curves exhibited by our model.[41]

5.5 Other Types of Firms

The critical feature in our story is that the investors bearing market risk cannot also be the ones doing the cross-sectional pricing of stocks, because those two activities imply very different levels of risk aversion. In this context, how should one think about firms such as investment banks with large investment portfolios of their own? They should be sophisticated institutions which are capable of active trading in individual stocks, yet they still sometimes bear significant exposure to market risk.

An investment bank with a proprietary trading portfolio can essentially be considered a closed-end fund. It actively trades individual stocks and the trading profits are equally distributed among shareholders. The costs of such a trading operation are reflected in the expenses of the firm and they are also equally distributed among shareholders, acting like a percentage fee on assets under management. In a competitive equilibrium we would expect the firm to raise capital by issuing shares until the abnormal return on the capital is approximately equal to the firm's costs. Hence, it makes no difference for our model

---

[41] Introducing heterogeneity into the beliefs of investors would not make transaction costs a more plausible explanation. 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 (far beyond anything we observe in the real market).

Similarly, any attempt to obtain large price effects from investor disagreement alone (e.g. trying to calibrate the model of Fama and French (2004) for this purpose) will face the same issue of counterfactually large short interest in individual stocks.

whether the active managers run open-end mutual funds, closed-end mutual funds, or public corporations with proprietary trading portfolios.[42]

However, it remains a puzzle why such an investment firm would simultaneously choose a very large exposure to market risk and very small exposure to idiosyncratic risk. This apparently schizophrenic attitude toward risk could result from benchmarking – just like an actively managed mutual fund, the investment firm can ignore market risk and let the end investors choose their own exposure to it. In the presence of short-sales costs, it may indeed be optimal for active managers to combine their long-short equity portfolios with large positions in the market portfolio.

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 fundamental puzzle in finance.

In this paper we propose an explanation for the puzzle. In traditional representative agent models it is implicitly 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 be an innocuous assumption. When most of the informed active investors are professional money managers who do not own the wealth they invest, the slope of the demand curve for a stock is determined by how much wealth they are given to manage. Since the active managers charge a fee for their services, the amount of wealth they manage and hence the slopes of demand curves are determined almost entirely by the fee and not by anyone's 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

[42] Berk and Stanton (2007) also present a related theoretical treatment of closed-end funds with rational capital allocation by end investors.

“inefficiency” which allows the active managers to recover what are presumably their fixed costs for acquiring information and actively trading on it. This severs the link between risk aversion and the demand curves for individual stocks. In contrast, the risk premium on the aggregate market portfolio is still entirely set by the end investors’ risk aversion since the broad asset allocation decision between stocks and bonds is a decision they make directly.

The magnitude of this effect can be surprisingly large. In our calibration, increasing the annual fee from zero, which corresponds to the CAPM benchmark, to 1.5% can increase the slope of the demand curve by a factor of over 500. 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 to an economically significant 3%. When we allow active managers to hold market risk and be benchmarked against it, as in the real money management industry, the price impact increases to 14%.

We believe this paper makes two main contributions. It suggests a generally applicable explanation to the persistent puzzle about downward-sloping demand curves, producing not only the correct sign for the effect but also the correct order of magnitude. More broadly, it provides a concrete illustration that the presence of financial institutions does have pricing implications, even without agency issues, broadening the conclusions of Ross (1989) and Allen (2001) about the relevance of institutions in asset pricing.

References

[1] Admati, A., 1985, "A Noisy Rational Expectations Equilibrium for Multiple Asset Securities Markets," *Econometrica*, vol. 53, no. 3, 629-657.

[2] Allen, F., 2001, "Do Financial Institutions Matter?" *Journal of Finance*, vol. 56, no. 4, 1165-1175.

[3] Allen, F. and D. Gale, 1994, "Limited Market Participation and Volatility of Asset Prices," *American Economic Review*, vol. 84, no. 4, 933-955.

[4] Amihud, Y. and H. Mendelson, 1986, "Asset Pricing and the Bid-Ask Spread," *Journal of Financial Economics*, vol. 17, no. 2, 223-249.

[5] Barberis, N. and A. Shleifer, 2002, "Style Investing," *Journal of Financial Economics*, vol. 68, no. 2, 161–199.

[6] Barberis, N., A. Shleifer, and J. Wurgler, 2004, "Comovement," *Journal of Financial Economics*, forthcoming.

[7] Barberis, N. and R. Thaler, 2003, "A Survey of Behavioral Finance," *Handbook of the Economics of Finance.*

[8] Berk, J.B., 1995, "A Critique of Size-Related Anomalies," *Review of Financial Studies*, vol. 8, no. 2, 275-286.

[9] Berk, J.B. and R.C. Green, 2004, "Mutual Fund Flows and Performance in Rational Markets," *Journal of Political Economy*, vol. 112, no. 6, 1269-1295.

[10] Berk, J.B. and R. Stanton, 2007, "Managerial Ability, Compensation, and the Closed-End Fund Discount," *Journal of Finance*, forthcoming.

[11] Biais, B., P. Bossaerts, and C. Spatt, 2006, "Equilibrium Asset Pricing and Portfolio Choice under Asymmetric Information," working paper.

[12] Black, F., 1986, "Noise," *Journal of Finance*, vol. 41, no. 3, 529-543.

[13] Blume, M.E. and R.M. Edelen, 2001, "On S&P 500 Index Replication Strategies," working paper, University of Pennsylvania.

[14] Chen, H., G. Noronha, and V. Singal, 2004, "The Price Response to S&P 500 Index Additions and Deletions: Evidence of Asymmetry and a New Explanation," *Journal of Finance*, vol. 59, vol. 4, 1901–1929.

[15] Chen, J., H. Hong, M. Huang, and J.D. Kubik, 2004, "Does Fund Size Erode Performance? Organizational Diseconomies and Active Money Management," *American Economic Review*, forthcoming.

[16] Chen, J., H. Hong, and J.C. Stein, 2002, "Breadth of Ownership and Stock Returns," *Journal of Financial Economics*, vol. 66, no. 2-3, 171-205.

[17] Chordia, T., 2001, "Liquidity and Returns: The Impact of Inclusion into the S&P 500 Index," working paper.

[18] Cohen, R., P. Gompers, and T. Vuolteenaho, 2002, "Who Underreacts to Cash-Flow News? Evidence from Trading between Individuals and Institutions," *Journal of Financial Economics*, vol. 66, no. 2-3, 409-462.

[19] Cremers, M. and A. Petajisto, 2007, "How Active Is Your Fund Manager? A New Measure That Predicts Performance," Yale ICF working paper.

[20] Daniel, K., M. Grinblatt, S. Titman, and R. Wermers, 1997, "Measuring Mutual Fund Performance with Characteristic-Based Benchmarks," *Journal of Finance*, vol. 52, no. 3, 1035-1058.

[21] Daniel, K., D. Hirshleifer, and A. Subrahmanyam, 2001, "Overconfidence, Arbitrage, and Equilibrium Asset Pricing," *Journal of Finance*, vol. 56, no. 3, 921-965.

[22] D'Avolio, G., 2002, "The Market for Borrowing Stock," *Journal of Financial Economics*, vol. 66, no. 2-3, 271-306

[23] De Bondt, W.F.M. and R. Thaler, 1985, "Does the Stock Market Overreact?" *Journal of Finance*, vol. 40, 793-805.

[24] De Long, J.B., A. Shleifer, L.H. Summers, and R.J. Waldmann, 1990, "Noise Trader Risk in Financial Markets," *Journal of Political Economy*, vol. 98, no. 4, 703-738.

[25] Denis, D.K., J.J. McConnell, A.V. Ovtchinnikov, and Y. Yu, 2003, "S&P 500 Index Additions and Earnings Expectations," *Journal of Finance*, vol. 58, no. 5, 1821-1840.

[26] Edmister, R., A.S. Graham, and W.L. Pirie, 1994, "Excess Returns of Index Replacement Stocks: Evidence of Liquidity and Substitutability," *Journal of Financial Research*, vol. 17, no. 3, 333-346.

[27] Elton, E.J., M.J. Gruber, and C.R. Blake, 2003, "Incentive Fees and Mutual Funds," *Journal of Finance*, vol. 58, no. 2, 779-804.

[28] Fama, E.F. and K.R. French, 2004, "Disagreement, Tastes, and Asset Prices," working paper, University of Chicago.

[29] Glosten L.R. and P.R. Milgrom, 1985, "Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, vol. 14, no. 1, 71-100.

[30] Gomes, J., L. Kogan, and L. Zhang, 2003, "Equilibrium Cross Section of Returns," *Journal of Political Economy*, vol. 111, no. 4, 693-732.

[31] Gompers, P.A. and A. Metrick, 2001, "Institutional Investors and Equity Prices," *Quarterly Journal of Economics*, 229-259.

[32] Greenwood, R., 2004, "Short- and Long-Term Demand Curves for Stocks: Theory and Evidence on the Dynamics of Arbitrage," *Journal of Financial Economics*, forthcoming.

[33] Grossman, S. and J.E. Stiglitz, 1980, "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, vol. 70, 393-408.

[34] Grossman, S. and M. Miller, 1988, "Liquidity and Market Structure," *Journal of Finance*, vol. 43, no. 3, 617-637.

[35] Gruber, M.J., 1996, "Another Puzzle: The Growth in Actively Managed Mutual Funds," *Journal of Finance*, vol. 51, no. 3, 783-810.

[36] Harris, L. and E. Gurel, 1986, "Price and Volume Effects Associated with Changes in the S&P 500 List: New Evidence for the Existence of Price Pressure," *Journal of Finance*, vol. 41, no. 4, 815-829.

[37] Holthausen, R.W., R.W. Leftwich, and D. Mayers, 1987, "The Effect of Large Block Transactions on Security Prices: A Cross-Sectional Analysis," *Journal of Financial Economics*, vol. 19, no. 2, 237-267.

[38] Holthausen, R.W., R.W. Leftwich, and D. Mayers, 1990, "Large-Block Transactions, the Speed of Response, and Temporary and Permanent Stock-Price Effects," *Journal of Financial Economics*, vol. 26, no. 1, 71-95.

[39] Hortacsu, A. and C. Syverson, 2004, "Search Costs, Product Differentiation, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds," *Quarterly Journal of Economics*, vol. 119, no. 2, 403-456.

[40] Jegadeesh, N. and S. Titman, 2001, "Profitability of Momentum Strategies: An Evaluation of Alternative Explanations," *Journal of Finance*, vol. 56, no. 2, 699-720.

[41] Kacperczyk, M.T., C. Sialm, and L. Zheng, 2004, "On Industry Concentration of Actively Managed Funds," *Journal of Finance*, forthcoming.

[42] Kaul, A., V. Mehrotra, and R. Morck, 2000, "Demand Curves for Stocks *Do* Slope Down: New Evidence from an Index Weights Adjustment," *Journal of Finance*, vol. 55, no. 2, 893-912.

[43] Kyle, A.S., 1985, "Continuous Auctions and Insider Trading," *Econometrica*, vol. 53, no. 6, 1315-1335.

[44] Loderer, C., J.W. Cooney, and L.D. van Drunen, 1991, "The Price Elasticity of Demand for Common Stock," *Journal of Finance*, vol. 46, no. 2, 621-651.

[45] Lynch, A. and R. Mendenhall, 1997, "New Evidence on Stock Price Effects Associated with Changes in the S&P 500 Index," *Journal of Business*, vol. 70, no. 3, 351-383.

[46] Mamaysky, H. and M. Spiegel, 2001, "A Theory of Mutual Funds: Optimal Fund Objectives and Industry Organization," working paper, Yale School of Management.

[47] Merton, R.C., 1987, "A Simple Model of Capital Market Equilibrium with Incomplete Information," *Journal of Finance*, vol. 42, no. 3, 483-510.

[48] Merton, R.C. and Z. Bodie, 2002, "The Design of Financial Systems: Towards a Synthesis of Function and Structure," working paper.

[49] Mitchell, M., T. Pulvino, and E. Stafford, 2004, "Price Pressure around Mergers," *Journal of Finance*, vol. 59, no. 1, 31-63.

[50] Miller, R.M., 2007, "Measuring the True Cost of Active Management by Mutual Funds," *Journal of Investment Management*, vol. 5, no. 1, 1-21.

[51] Morck, R. and F. Yang, 2001, "The Mysterious Growing Value of S&P 500 Membership," NBER working paper.

[52] Nanda, V., M.P. Narayanan, V.A. Warther, 2000, "Liquidity, Investment Ability, and Mutual Fund Structure," *Journal of Financial Economics*, vol. 57, 417-443.

[53] Petajisto, A., 2004, "The Index Premium and Its Implications for Index Funds," Yale ICF working paper.

[54] Rashes, M.S., 2001, "Massively Confused Investors Making Conspicuously Ignorant Choices (MCI–MCIC)," *Journal of Finance*, vol. 56, no. 5, 1911–1927.

[55] Ross, S., 1976, "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, vol. 13, 341-360.

[56] Ross, S., 1989, "Institutional Markets, Financial Marketing, and Financial Innovation," *Journal of Finance*, vol. 44, no. 3, 541-556.

[57] Scholes, M., 1972, "The Market for Securities: Substitution versus Price Pressure and the Effects of Information on Share Price," *Journal of Business*, vol. 45, 179-211.

[58] Sharpe, W.F., 1981, "Decentralized Investment Management," *Journal of Finance*, vol. 36, no. 2, 217-234.

[59] Shleifer, A., 1986, "Do Demand Curves for Stocks Slope Down?" *Journal of Finance*, vol. 41, no. 3, 579-590.

[60] Shleifer, A. and R.W. Vishny, 1997, "The Limits of Arbitrage," *Journal of Finance*, vol. 52, no. 1, 35-55.

[61] Treynor, J.L. and F. Black, 1973, "How to Use Security Analysis to Improve Portfolio Selection," *Journal of Business*, vol. 46, no. 1, 66-86.

[62] Van Nieuwerburgh, S. and L. Veldkamp, 2006, "Information Acquisition and Portfolio Under-Diversification," working paper.

[63] Wermers, R., 2000, "Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses," *Journal of Finance*, vol. 55, no. 4, 1655-1695.

[64] Wurgler, J. and K. Zhuravskaya, 2002, "Does Arbitrage Flatten Demand Curves for Stocks?" *Journal of Business*, vol. 75, no. 4, 583-608.