Informative Prices and the Cost of Capital Markets

Working Paper

Robert Turley*

November 2012

*Harvard University Department of Economics and Harvard Business School. Email: turley@fas.harvard.edu. Mailing address: Baker Library 244B, Boston MA 02163. I thank Thomas Philippon for sharing data and advice. I am also grateful for insightful comments from John Campbell, Josh Coval, Alp Simsek, Robin Greenwood and Eduardo Davila.

Abstract

We spend an enormous amount of resources actively investing in financial markets, a cost which has increased dramatically over the past few decades. Using historical data from US equity markets, I document the connection between trading efficiency, market activity, and the information content of asset prices. These relationships are predicted by a stylized model, where the increasing efficiency of financial transactions leads to more—not less—spending on financial activity. This effect grows stronger as the investment horizon contracts. To identify the importance of this proposed efficiency explanation, I use the natural experiment that occurred when the SEC implemented Rule 19-b in May of 1975, finding strong evidence that transaction efficiency is an important driver of the modern increase in the cost of capital markets.

JEL classification: E44, G2, G12

1 Introduction

Investors spend a great deal of time and money speculating on financial valuations or hiring others to trade on their behalf. While criticizing speculation is always fashionable, the scale of the recent increase in resources spent on capital markets has many people concerned that we are wasting talent and resources. There seems to be little consensus among financial economists regarding the value of this speculative activity; however, it is easy to observe the increase in quantity. Historically, the share of national income spent on financial market activity remained relatively stable until the mid-1970s, when the financial sector began to grow much more rapidly than the aggregate US economy. Before rushing to judge whether we now spend too much, or too little, on active investing, we need theory and evidence that promise to explain the root cause of this growth and the resulting effect on asset prices.

In this paper, I document how the sharp decline in the cost of financial transactions facilitated the modern increase in financial activity. To clarify the forces at work, I present a stylized model of an economy with a financial sector that allows investors to trade ownership claims on a risky investment. The supply of investment responds to asset prices, and investor demand drives costly financial activity. Investors decide how much of their resources to employ researching the future prospects of the uncertain outcome, and market transaction costs affect the quantity and time horizon of informed speculation. We see the surprising result that the financial sector consumes more resources through spending on active investing as it operates more efficiently. As dynamic trading strategies become feasible, the model suggests that the information content of asset prices increases, especially over short-horizons.

Historical data on US market activity and asset prices confirm these predictions. The most significant decrease in transaction costs occurred in 1975, when on May Day the SEC demanded that stock exchanges end the practice of forcing a fixed commission schedule on all equity transactions. In response to broker competition, the average cost of institutional trading plummeted to about half of previous levels.[1] This event is significant not only in the historical time series, but it also provides a natural setting for identifying the causal mechanism. This regulatory change leads to

[1] US Securities and Exchange Commission, Directorate of Economic and Policy Research. *Staff Report on the Securities Industry in 1978* (July 26, 1979)

a surge in capital market spending, trading, and compensation, with an impact that predictably varies across investment characteristics and time horizons.

The efficiency of modern financial markets enables dynamic trading strategies and encourages investors to spend more resources on research and trading, but increased efficiency does not necessarily align the incentives of private speculators toward activities with the greatest social benefit. Returning again to the stylized model shows that increases in the efficiency of financial market operations may lead to less efficient economic outcomes.

1.1 Spending on Capital Market Activity

Consider how much the United States spends on capital market activities each year as a share of total national production. Figure 1 shows the cost of capital markets as a percentage of the GDP of the US private sector, where capital market spending consists of the profits and employee compensation tabulated using the gross value added measures reported by United States Bureau of Economic Analysis (BEA).[2] The cost of capital markets is remarkably stable for approximately half a century. Beginning with a cost of 0.27% of GDP in 1920 to a cost of 0.35% in 1970, spending stays fairly close to its average of 0.32% with the exception of a moderate dip around World War II. Then, a little before 1980, we notice a dramatic surge in the cost of capital markets to the point where capital markets now consume two percent of annual spending.

Philippon (2012) lays out the scope of the historical challenge as he tabulates the costs and quantities of various financial activities over the past 130 years in the United States. In his analysis, it appears that the unit cost of financial intermediation has remained relatively stable over time despite advancements in technology. He notes a puzzling increase in the cost of financial activity over the past 30 years that he cannot explain with a corresponding increase in the quantity or quality of financial services.

With a particular focus on this modern period, Greenwood and Scharfstein (2012) attribute the modern growth of the financial sector as a whole to two specific components: an increase in active investing and an expansion in credit markets. To contrast these two culprits, I allocate the corresponding financial activities from the national industry accounts data, as shown in Table 1. The resources consumed in credit and banking activities grew significantly over the past century

[2] A complete description of the underlying data will be available in an online appendix.

but follow a distinct pattern from the resources spent investing in financial markets. The upper plot in Figure 2 shows both activities consumed a growing fraction of GDP, but the cost of banking and credit expanded at steady consistent pace since World War II while the surge in trading and investing seems to be a more recent phenomenon. Unlike the capital markets sector, the lower plot of Figure 2 shows the historical compensation of employees in the banking and credit sector differs only slightly from the private sector average and increases only moderately in recent decades.

1.2 Theories of financial investment distortions

Dissatisfaction with the quantity of talent and resources consumed by financial markets seems to peak during economic downturns. Amidst the Great Depression, Keynes criticized American financial markets, arguing, "when the capital development of a country becomes the by-product of the activities of a casino, the job is likely to be ill-done."[3] On the other hand, the broad impact of financial crises could also suggest we need a large and highly compensated financial sector to replace animal spirits with dispassionate analysts.

Certainly, there is a need to understand the circumstances and incentives that pull resources toward financial markets. What gives rise to a distorted financial sector? Economic research offers three explanations for outsized financial activity: irrational investors do not know they trade too much, rational investors cannot help trading too much, or perhaps the industry is rife with rent-seeking.

Financial markets seem to be amazingly adroit at exploiting irrational beliefs and behaviors. Fanciful trading or the decision to pay exorbitant fees to popular investment managers may funnel unnecessary fees into finance and have other negative consequences (De Long, Shleifer, Summers and Waldmann, 1989).

In a model where market participants are assumed to be rational, they may still spend too much on active investment because inference is difficult (Pástor and Stambaugh, 2010) or out of a desire to avoid being the greater fool when negotiating transactions. Glode, Green and Lowery (2012) present this situation as an arms race externality for financial expertise. The model presented by Bolton, Santos and Scheinkman (2011) has a similar mechanism; opaque markets attract talent

[3]Keynes, John Maynard, *The General Theory of Employment, Interest and Money* (London: Macmillan, 1936), page 159.

and more informed valuations lure the best investments away from public exchanges.

These explanations capture important aspects of financial markets, but neither seems uniquely modern. If traders are foolish now, they were foolish before. Shrewd traders will always prefer to be better informed than their counterparty. We are forced to ask: what changed?

Philippon and Reshef (2013) point toward the rent-seeking channel, and propose the growth in compensation is a result of deregulation. The active government oversight intended to curb the worst excesses in the financial markets of the 1920s was gradually relaxed 50 years later, and Philippon and Reshef propose rents lured talent from more productive endeavors (Murphy, Shleifer and Vishny, 1991).

Supporting this view, Bai, Philippon and Savov (2012) suggest modern asset prices show no increase in their information content over the past 50 years. They suggest the increase in financial spending may result from rent extraction, suggesting the growth in active investment has had little effect on asset prices.

1.3 Understanding the causes and consequences of the cost of capital markets

With so much highly compensated talent flowing into investment management, it is hard to believe that asset prices are no more informative in the modern information age than they were in the bygone era when investors in top hats exchanged small pieces of paper. As an alternative explanation for the root cause of the modern growth of capital markets, I propose technological efficiency. The decreasing cost of transacting makes dynamic trading strategies feasible and draws talent and technology toward acquiring faster paced information. Confirming the results of Bai et al. (2012), I find only very weak evidence that modern asset prices capture more long-horizon information; however, I find strong evidence of an increase in active trading and information content at horizons of less than one year.

To help frame the empirical findings, I present a stylized model illustrating the role of trading horizons in costly capital markets. The key comparative static will measure the effect of increases in trading efficiency. The model predicts that as the cost of financial activity decreases, total spending in the financial sector actually increases, especially for short-horizon speculation.

This explanation has a large degree of empirical success in explaining aggregate spending on capital markets over time, particularly in regard to aggregate spending on active investing (French, 2008). More efficient transaction costs lead to higher quantities of informed trading, providing an underlying explanation for Greenwood and Scharfstein's observation that the observed growth of modern finance coincides with a growth in actively investing. The events of May 1975 highlight the significance of this mechanism, as the SEC instituted rule 19b and replaced the high trading commissions enforced by stock exchange members with competitive transaction rates. Using this event and information from historical fee schedules, we observe how the operational efficiency of capital markets affects the financial industry and market prices.

This paper provides new evidence on the changes that caused and accompanied the modern growth in the cost of capital markets. Linking these findings to economic theory clarifies the underlying incentives and opens the door to the broader question of whether the returns to finance are worth the cost.

2 A Stylized Model of Capital Markets

In this section, I present a stylized model of capital markets where the supply of the risky investment responds to asset prices and where the financial market is costly to operate. I will show how changes in the cost of transacting affect the quantity of resources spent on finance and affect the characteristics of asset prices.

To better understand the role financial markets play, consider an illustrative, general equilibrium framework where investors spend resources in acquiring information and engaging in costly transactions. In the spirit of the Q-theory (Brainard and Tobin, 1977), the supply of investment will respond to the market price, so the information in asset prices plays a key role in capital allocation. Ultimately, we want to observe how changes in the cost of transacting affects the resources spent in capital markets. Additionally, the model will distinguish between short-run and long-run behavior, generating novel predictions relating the growth in capital market spending to asset prices which will be confirmed in the data.

Unlike the opaque bilateral setting of Glode et al. (2012), all market prices in the model will be publicly observed, which has historically been true for equity markets and is becoming increasingly common across asset classes. The setup more closely resembles the endogenous information setting of Grossman and Stiglitz (1980), adding the salient features necessary to model a costly financial market and multiple time horizons.

The key comparative statics will be the impact of an exogenous change of transaction costs on total capital market spending and the information content of asset prices, noting the differential impact by trading horizon. I briefly mention the welfare implications in section 5.

2.1 The Setting

The supply of risky investment

Consider a risky investment traded publicly over a T periods ($t \in [1, T]$) prior to yielding an uncertain payout X consumer in period $T+1$, where the uncertain component of X is

$$X - \mathrm{E}[X] = \sum_{t=1}^{T} \theta_t + \varepsilon. \tag{1}$$

Each of the component random variables are independent, mean-zero, and normally distributed with variances σ_θ^2 and σ_ε^2. The full, random component $\sum \theta_t$ becomes public knowledge in period $T+1$. However, market participants can spend resources to discover the information in period 0, and they will be termed long-horizon investors. Alternately, short-horizon investors may spend a smaller amount of resources to discover each piece of short horizon information (θ_t) in period t. The random component ε cannot be observed prior to period $T+1$.

The quantity of the risky investment is responsive to investment demand, allowing the quantity of shares in one period, Q_t, to increase or decrease with the market price, P_t. For simplicity, we'll model this as a linear supply curve, with slope parameter $b > 0$. The change in investment supply will be

$$Q_{t+1} - Q_t = b\left(P_t - P_{t-1}\right). \tag{2}$$

where the initial price is assumed to be the unconditional expectation, $P_0 = \mathrm{E}_0[P_1]$. By construction, the supply of investment is fixed in the short-run (contemporaneous with the trading period) and responds to financial market prices over longer horizons (the next period).

Investors and financial markets

The agents will be modeled by a continuum of identical investors. Wealth can be transferred across periods at an interest rate of zero and is consumed in the final period. Each investor is endowed with w_0 units of wealth (measured in units of final consumption) and a share, q_0, of the risky investment. By construction, the total initial quantity of investment is $Q_0 = \int_{i\in[0,1]} q_{0,i}\mathrm{d}i$.

Individuals can choose whether they want to acquire information and actively speculate based on the difference between their valuation and the observed market price. To learn the full value of $\sum\theta_t$ during the first trading period requires paying k_L, whereas short-horizon traders who only learn each component θ_t at time t pay $k_S \leq k_L$. Alternately, investors may choose to infer their valuations from the public market price. Since their valuations will not differ from the market price, they will not actively trade and I'll refer to these traders as passive, though they might make trades driven by changes in their uncertainty.

Each individual seeks to maximize expected CARA utility of final consumption. For convenience, we'll denote the consumption of investor i as their final wealth, w_i, with associated expected utility $\mathrm{E}[-\exp\{-aw_i\}]$ for absolute risk aversion parameter a.

Investors must commit whether to spend resources on information in period $t = 0$ before any trading happens. In subsequent periods prior to the final outcome, investors may choose to trade their holdings of the risky asset at the prevailing market price. The transaction costs associated with capital markets are passed directly through to investors. For analytical convenience, we'll assume they take a quadratic form so that the trading from a prior holding of $q_{i,t-1}$ shares in period $t-1$ to $q_{i,t}$ during the trading in period t will result in a transaction cost of $\frac{c}{2}\left(q_t - q_{t-1}\right)^2$.

We can describe the evolution of investor wealth as

$$w_{i,t+1} = w_{i,t} + q_{i,t}\left(P_{t+1} - P_t\right) - \frac{c}{2}\left(q_{i,t+1} - q_{i,t}\right)^2 \tag{3}$$

where agents are identically endowed with w_0 consumption and q_0 shares of the risky investment. In the final period, the price of the risky investment will simply be the outcome, i.e. $P_{T+1} = X$.

Portfolio choice

The linear-CARA-normal framework allows the expected utility from the perspective of investor i in trading period t to be calculated as

$$\mathrm{E}_{i,t}\left[-\exp\left\{-aw_i\right\}\right] = -\exp\left\{-a\mathrm{E}_{i,t}\left[w_i\right] + \frac{a^2}{2}\mathrm{Var}_{i,t}\left[w_i\right]\right\}. \tag{4}$$

Through monotonic transformations, the investor can maximize the certainty-equivalent, which takes the mean-variance form, $\mathrm{E}_{i,t}\left[w_i\right] - \frac{a}{2}\mathrm{Var}_{i,t}\left[w_i\right]$. The concavity of the problem suggests we can find the optimal portfolio in each period, $q^*_{i,t}$, at the point where the first order condition holds, $\frac{\partial}{\partial q_{i,t}}\mathrm{E}_{i,t}\left[w_{i,3}\right] = \frac{a}{2}\frac{\partial}{\partial q_{i,t}}\mathrm{Var}_{i,t}\left[w_{i,3}\right]$.

To motivate the optimal portfolio rules, we can work backwards from the final trading period. The optimal portfolio $q^*_{i,T}$ in last trading period that maximizes the utility of consumption in the subsequent period will have the associated first order condition

$$q^*_{i,T} = \frac{\mathrm{E}_{i,T}\left[X - P_T\right] + cq_{i,T-1}}{a\mathrm{Var}_{i,T}\left[X\right] + c}. \tag{5}$$

This is the classic myopic portfolio rule with a transaction cost adjustment. In the numerator, we see the optimal portfolio increases linearly with the expected return, $\mathrm{E}_{i,T}\left[X - P_T\right]$. The second term in the numerator shows how much transaction costs discourage trading by anchoring the portfolio at the initial position, $q_{i,T-1}$. The magnitude of the transaction costs, c, determines the extent to which this affects the optimal portfolio.

In solving the model, I will show how the anchoring feature of transaction costs results in optimal portfolio rules that are a weighted average of their myopic, one-period expected return and the returns offered in future periods.

2.2 Equilibrium

In this setting, investors can be grouped into three types based on their information sets. The mass of agents of type j are those who pay k_j for their investment information will be measured as the quantity $\lambda_j \in [0, 1]$.

Definition In a rational expectations equilibrium,

(a) markets will clear

(b) investors will choose to spend resources on information to maximize ex ante utility, leading to an allocation $\{\lambda_L, \lambda_S\}$ and where $\lambda_N = 1 - \lambda_L - \lambda_S$ is the fraction of individuals who will only infer information from market prices

(c) investors of each type have an optimal demand function $q_{i,t}(P_t)$ for the risky asset conditional on the market price, which will be constructed from their rational beliefs about random variables (θ_t and ν_t) conditional on the observed price.

Market clearing

It will be useful to explicitly define market clearing. Noisy supply shocks will add uncertainty so that the market price does not perfectly reveal all information. Specifically, the total quantity of investment supply will equal investment demand,

$$Q_t = \sum_i \lambda_i q_{i,t} + \frac{\nu_t}{a\sigma_\varepsilon^2 + c}, \tag{6}$$

comprising the sum of the individual demands ($q_{i,t}$) times the mass of the investor type (λ_i) plus the scaled demand shock $\nu_t \sim N\left(\sigma_\nu^2\right)$. The values in the denominator scale the shock by variance and transaction costs. In this sense, the noise can be interpreted in the same way as the demands of an informed investor, as can be seen from demand function (5), but obviously the shock is unrelated to the actual final payout of the investment.

Intuition

To build the intuition behind this model and its equilibrium, consider Figure 3. For this particular illustration, this will assume just one trading period ($T = 1$) and there is no distinction between long-horizon and short-horizon informed investors, though the paper will generally consider $T > 1$ in order to highlight the importance of time horizon. The left panel plots the fraction of informed speculators along the horizontal axis, ranging from 0 to 1. The vertical axis measures expected utility for both the informed speculators and the expected utility for the uninformed, passive investors. When there are no informed speculators, the information advantage is obvious as the expected utility for informed active investors is significantly higher than that of the passive investors who observe only the market price. As the fraction of the informed investors increases, the difference between the two expected utilities decreases. This is the general case, and the intuition extends to the multiple period setting; as the market price becomes more informative the relative advantage of paying for the information decreases. With these parameters, the equilibrium point of indifference between acquiring the costly information occurs at the point where approximately $\frac{1}{4}$ of the investors acquire the costly information. To the right of the equilibrium point, the trading profits resulting from learning more about the risky outcome θ are not worth the resources it could cost (k).

On the right panel, the horizontal axis continues to measure the fraction of informed speculators, and on the vertical axis we see the equilibrium price. In the case of no informed investors, the variation in price is entirely due to the supply shocks ν. As the fraction of informed traders increases, we see two effects. The average price increases as investors are willing to commit more capital to investment because there is less uncertainty. Additionally, the variance of the market prices increases. This is because the price now also contains information about the investment prospects. Not surprisingly, the information content of asset prices levels of around the equilibrium point, further evidence that little additional value is gained acquiring information that is already largely in the market price.

Proposition 1 (Equilibria)**.** *There exist rational expectations equilibria under the assumed parameter restrictions* ($0 < k_S < k_L$)*.*

The proof for the one-period case ($T = 1$) should be clear from the discussion above. There will be no long-horizon traders. Since the expected utilities are continuous in $\lambda \in [0, 1]$, we simply need to appeal to the intermediate value theorem for existence. The difference between the expected utility of the informed and uninformed traders will nearly always be monotonically decreasing in λ, which guarantees uniqueness.

The same intermediate value approach guarantees a unique solution in the case of multiple periods ($T > 2$) in the case where one or more type is always inferior and has optimal weight zero. The existence of the multiple horizon solution when there is a positive mass of each of the three types can be motivated by working backwards from the final period. In the final period, informed traders face a situation identical to the one-period model. In prior periods, the relative advantage to the long-horizon information is decreasing in λ_L. The mass of investors in λ_S will be uninformed about the information θ_{t+k} (for $k > 1$), and like the uninformed investors, can infer more information as λ_L increases. As long as there are positive quantities of each investor type, the marginal effect of more traders will follow the same relative rank impact on ex ante utility, guaranteeing a unique solution.

2.3 Characterizing a multiple horizon solution ($T = 2$)

To characterize the analytical differences between long-horizon and short-horizon speculation, I will more fully characterize the solution for $T = 2$. In this setting, the outcome will be a long-run event in the first period and a short-run event in the second period, which immediately precedes the investment outcome. After this short-horizon trading is complete, investor i will consume

$$w_i = w_0 + q_0 P_1 + q_{i,1}\left(P_2 - P_1\right) + q_{i,2}\left(X - P_2\right) - \frac{c}{2}\left(\left(q_{i,1} - q_0\right)^2 + \left(q_{i,2} - q_{i,1}\right)^2\right) - k_i. \quad (7)$$

Assuming linearity and the resulting expectations

To calculate the investor demand functions, we need to know their expectations, which will be affected by the information they perceive from the market prices they observe. I will assert and then prove that the market prices can be expressed as linear functions of the unknown variables,

$$P_1 = \bar{P}_1 + \beta_1\theta_1 + \beta_2\theta_2 + \beta_{\nu_1}\nu_1 \quad (8)$$

and

$$P_2 = \bar{P}_2 + \beta_P\left(P_1 - \bar{P}\right) + \beta_3\theta_1 + \beta_4\theta_2 + \beta_{\nu_2}\nu_2. \quad (9)$$

The unknown coefficients are derived in the appendix, thus confirming the assumed linear functional form.

Additionally, to help with the notation and intuition, we note that the beliefs of uninformed and short-run traders hold about X from observing the market price in period 1 will be affected by the variation in price. We can express these expectations as

$$\mathrm{E}_{S,1}\left[X\right] = \bar{X} + \rho_{S,1} Y_{S,1} \tag{10}$$

where

$$\begin{aligned} Y_{S,1} &= \theta_2 + \frac{\beta_{\nu_1}}{\beta_2}\nu_1 && (11) \\ &\propto \left(P_1 - \bar{P}_1 - \beta_1\theta_1\right) && (12) \end{aligned}$$

and so that $\rho \in [0,1]$ is a simple function of the assumed parameters

$$\rho = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \left(\frac{\beta_{\nu_1,1}}{\beta_{\theta,1}}\right)^2 \sigma_\nu^2}.$$

The investors who have spent no resources on information simply take valuations from their deviation from the market price

$$\left(\mathrm{E}_{N,1}\left[X\right] - \bar{X}\right) \propto \left(P_1 - \bar{P}\right) \tag{13}$$

Portfolio optimization in period 2

The investors will be categorized by the trading period in which they receive information about θ: in the long-horizon (L), short-horizon (S) and not at all (N).

For each of the three investor types (L, S, and N), we can express their optimal portfolio in terms of their prior position and their current expectations $\mathrm{E}_{i,2}\left[X\right]$ and $\mathrm{Var}_{i,2}\left[X\right]$. The long-run and short-run speculators will both know θ_1 and θ_2 in period 2 so $\mathrm{E}_{L,2}\left[X\right] = \mathrm{E}_{S,2}\left[X\right]$. The associated variance will be $\mathrm{Var}_{L,2}\left[X\right] = \mathrm{Var}_{S,2}\left[X\right] = \sigma_\varepsilon^2$. From (5) we can conclude that the optimal portfolio for these two types of investors will be

$$q_{L,2}^{*} = \frac{\left(\bar{X} + \theta_1 + \theta_2 - P_2\right) + c q_{L,1}^{*}}{a\sigma_\varepsilon^2 + c} \tag{14}$$

and

$$q^*_{S,2} = \frac{\left(\bar{X} + \theta_1 + \theta_2 - P_2\right) + cq^*_{S,1}}{a\sigma^2_\varepsilon + c} \tag{15}$$

The optimal portfolio for the investors who purchase no information

$$q^*_{N,2} = \frac{\mathrm{E}_{N,2}\left[X - P_2\right] + cq^*_{N,1}}{a\mathrm{Var}_{N,2}\left[X\right] + c} \tag{16}$$

depends on the expectations, $\mathrm{E}_{n,2}[\theta]$ and $\mathrm{Var}_{n,2}[\theta]$, which will be derived later.

Portfolio optimization in period 1

When investing for the long-run (in period 1), investors choose their allocation aware of their optimal short-run portfolio rules in equations (14 - 16). Those short-run rules show that each portfolio allocation is linearly related to the expected return ($\mathrm{E}_i\left[X - P_2\right]$) and the prior portfolio allocation ($q_{i,1}$).

The form of the period 1 demand function for long-horizon investors is similar to that of the other two investor types. It is derived by substituting the period 1 demand from equation (14) into equation (7) and taking the first order conditions to find the optimal portfolio

$$q^*_{L,1} = \frac{\left(1 - \Gamma\right)\mathrm{E}_{L,1}\left[P_2 - P_1\right] + \Gamma\mathrm{E}_{L,1}\left[X - P_1\right] + cq_0}{\Omega + c\left(1 + \left(\frac{a\sigma^2_\varepsilon}{a\sigma^2_\varepsilon + c}\right)^2\right)} \tag{17}$$

where the tilt toward the long-run return is

$$\Gamma = \underbrace{\frac{c}{a\sigma^2_\varepsilon + c}}_{\text{return next period}} + \underbrace{a\frac{\left(2a\sigma^2_\varepsilon + c\right)\beta^2_{\nu_2}\sigma^2_\nu a^2\sigma^4_\varepsilon}{\left(a\sigma^2_\varepsilon + c\right)^4}}_{\text{prefer to avoid adverse } \nu_2}$$

and the variance

$$\Omega = \underbrace{\left(\frac{c}{a\sigma^2_\varepsilon + c}\right)^2\sigma^2_\varepsilon}_{\text{variance of } X} + \underbrace{\left(\frac{a\sigma^2_\varepsilon}{a\sigma^2_\varepsilon + c}\right)^4\beta^2_{\nu_2,2}\sigma^2_\nu}_{\text{variance in } P_2}.$$

To develop some intuition for this long-horizon portfolio rule in equation (17), consider the three terms in the numerator. As before, there is a weight pulling the optimal portfolio toward the initial position, q_0 as a result of transaction costs. The other two terms are a weighted average of the myopic expected return, $\mathrm{E}_{L,1}[P_2 - P_1]$ and the long-run expected return, $\mathrm{E}_{L,1}[X - P_1]$, with respective weights $(1 - \Gamma)$ and Γ.

The weight Γ that the investor tilts toward the long-horizon return will always be weakly positive, $\Gamma \in [0, 1)$, and its magnitude will increase with transaction costs. The relationship with transaction costs arises from the investor recognizing positions taken today will persist into the future due to the anchoring effect of transaction costs. Additionally, there is some uncertainty in the price next period, so investors have an incentive to lock in P_1 now rather than pay an uncertain P_2.

The demand functions for the short-run and uninformed investors take an identical form, with slightly different values for Γ and Ω. These can be found in the appendix.

Market Clearing and Investor expectations

The investors will form expectations about investment prospects (X) and the effect of the noise shocks (ν_1 and ν_2) from the market price. Intuitively, investor expectations of θ increase in the market price, but larger noise shocks dampens this relationship. Complete expressions for investor expectations can be found in the appendix, and I verify the assumed linear relationship between prices and the unknown variables as suggested in equations (8) and (9).

2.4 The impact of more efficient transactions

Let's now turn to the question of what happens if the financial sector is more operationally efficient and the cost of transacting decreases. I consider two key comparative statics: how does this affect total active investment management ($\frac{\partial \sum \lambda_i}{\partial c}$) and how does this effect differ by investment horizon ($\frac{\partial \lambda_s}{\partial c}$ versus $\frac{\partial \lambda_l}{\partial c}$).

Proposition 2 (More active management)**.** *As the cost of transacting decreases, total informed trading increases,*

$$\frac{\partial \sum \lambda_i}{\partial c} \leq 0$$

and this becomes a strict inequality if there is any interior solution (i.e. 0¡$\lambda_j < 1$ *for some* j*).*

The value gained from information lies in the ability capitalize on the information through active trading. Clearly, in the limiting case, $\lim_{c \to \infty} \lambda_n^* \to 1$. For interior solutions, we must consider the marginal impact of transaction costs on the relative utility of informed and uninformed investors. The unconditional expected utility of an informed speculator will be a decreasing, continuous function of transaction costs. The unconditional expected utility of a passive investors will also decrease–but much less rapidly. Hence, $\frac{\partial \sum \lambda_N}{\partial c} \geq 0$. Since these functions are continuous, equality will only hold in the corner solutions where marginal changes in expected utility have no effect on the allocations of investor type.

Proposition 3 (Shorter investment horizons)**.** *Lower transaction costs have a greater effect on short-horizon investors than long-horizon investors,*

$$\frac{\partial \lambda_S}{\partial c} \leq \frac{\partial \lambda_L}{\partial c}$$

with strict inequality for interior solutions (i.e. $\lambda_L \in (0,1)$ *and* $\lambda_S \in (0,1)$*).*

This result comes from the fact that the short-horizon investors' optimal portfolio contains a subset of the information of the long-horizon investor. So the desire to spread trading over a longer horizon is offset by the fact that the short-horizon signal in period 1 (θ_1) may be in the opposite direction as the signal in period 2 (θ_2). As a result, short-horizon traders are forced to trade more for the same expected return.

In fact, in a model with many periods (T large), the short-horizon traders will find that the independence of θ_t makes trading in the earliest periods costly relative to the weakness of their accumulated signal. As the final horizon approaches, the short-horizon traders will be more inclined to trade as their accumulated signal is stronger and less likely to suggest they need to unwind their trades because of future information.

In contrast, the long-horizon traders are eager to trade on their information as early as possible, but they submit to spreading their trading across later periods in their desire to minimize their transaction costs. There are also information advantages to spreading out trades, since larger trades move prices and allow other traders to freely infer the costly information, but the infinitesimal traders do not absorb this externality.

3 Explaining the Empirical Growth in Capital Market Spending

A key contribution of this paper is document the relationship between the efficiency of financial transactions and the growth of modern finance. As improvements in technology and market organization make transactions less costly, we should expect to see the volume of transactions increase. This simply follows from the economic Law of Demand. A more surprising result is that as financial costs decrease, total spending on finance increases. This is fundamentally a statement about elasticities.

In this section, I focus on establishing the relationship between financial efficiency and the aggregate measures of financial spending and activity. I use timing to assert causality in the Granger sense, and using the (plausibly) exogenous historical break in May of 1975. The evidence is statistically strong but open to the criticism that the changes in efficiency may be interrelated with contemporaneous events. In section 4, I will use cross-sectional variation in the panel data to establish even stronger results and focus more explicitly on measuring the information content and investment horizon, two key features of the model.

3.1 A time series of transaction costs

With the possible exception of the very recent past, brokerage commissions were the primary cost in trading equities (Berkowitz, Logue and Noser, 1988). They funded all the operations required in financial market transactions. To test the efficiency explanation for the growth of capital markets, I construct a historical time series that measures the representative cost of transacting. The measure I propose splices two date ranges: 1927-1975 and 1975-2010.

Pre-1975: the NYSE fixed commission schedule

From its founding in 1792 up to 1975, the New York Stock Exchange (NYSE) enforced a minimum commission schedule on all of its member firms. The smaller, regional exchanges mirrored the commission schedule of the NYSE, and in the rare cases where they didn't, they faced enormous industry pressure to conform. The stated goal was to "prevent competition amongst the members" to protect their profits. Exchange members referenced the general fear of unfettered trading and defended high trading costs by observing that "a very low or competitive rate would also promote speculation." [4]

An example commission schedule, corresponding to the NYSE rates for 1956 is displayed in Figure 5. We can see how the formula defining the commission rate is a function of the nominal share price. Purchasing a round lot (100 shares) of a stock costing \$30 per share, for example, would have a commission of \$15 +0.5 times \$30. A round lot of a \$60 stock would cost \$35 +0.1 times \$60.

To construct a time series of the average transaction cost prior to 1975 I collect the NYSE commission schedules, including the NYSE annual fact books and the monthly S&P Stock Owners Guide. Combining these commission schedules with trading volume and price data from CRSP,[5] I construct an annual series of the weighted average cost of trading.

May Day 1975

In the aftermath of the financial disasters surrounding the Great Depression, the Securities Exchange Act of 1934 charged the Securities and Exchange Commission (SEC) with regulating and approving changes to any enforced commission schedules. Over the following forty years, the NYSE would periodically submit proposals to increase rates. A pattern emerged whereby the NYSE would complain about the rising costs and shrinking profits of its members, propose an increase in the commission schedule in order to maintain an appropriate level of profitability, and they would get immediate approval from the SEC.

In 1968, however the SEC scrutinized the latest proposed increase with more skepticism. Regulators asked why the cost of transacting in the financial markets could not itself be the product of a competitive response. The response from the exchange was emphatic: "One does not move the keystone of an industry which facilitates the raising of the bulk of new capital for this country...Negotiated rates would bring a degree of destructive competition."[6]

Although the SEC continued to approve a series of regular increases, this initial dissatisfaction was not placated. On January 23, 1975 the SEC adopted rule 19b, requiring all stock exchanges to end the practice of the fixed commission schedule and allow members to set rates competitively.

[4] Report of the Committee Appointed Pursuant to House Resolutions 429 and 504 to Investigate the Concentration of Control of Money and Credit, H.R. REP. NO. 62-1593

[5] Center for Research in Security Prices. Graduate School of Business, The University of Chicago (2012), Used with permission. All rights reserved.

[6] Richard Hack, NYSE president (August 19, 1968)

This rule was to go in effect on May 1, 1975. Distressed brokers and the popular press referred to the deadline as May Day.

As brokers competed for the first time on trading costs, there was a sharp drop in costs, especially for institutional investors. The SEC instituted a number of studies trying to measure the impact of their rule. Only two weeks after the beginning of competitive rates, the SEC Commissioner noted that they have seen sharp price cutting, in some instances to half or less of previously prevailing rates.[7] The SEC study of 1978 concluded that institutional trading costs had stabilized to a level 52.9% below their fixed rate levels.[8] Interestingly, the costs to individual traders decreased only moderately, giving rise to price discrimination among investor types (Tinic and West, 1980).

Post-1975: NYSE member financial statements

To continue the time series measuring the cost of transacting in the modern period of negotiated commissions post-1975, I collect commission revenues from the member financial statements of the NYSE and divide them by trading volume to estimate the weighted average cost per share.

Figure 6 shows the composite time series from 1927 to 2010. We can see the significant increase in the early 1930s followed by a relatively steady increase in costs for almost 50 years until the sudden drop resulting from the events of May 1975. To ensure the aggregate time series is a fair representation of aggregate transaction costs, I compare it to a number of independent measures. These include: the survey results from Greenwich Associates, a consultancy that surveys institutional investors regarding the costs they pay for their transactions; the SEC studies measuring transaction costs in the wake of rule 19-b; and for historical purposes, the cost associated with trading a $30 stock, holding the nominal share price constant through the duration of the fixed commission schedule. Each of these measures corresponds relatively closely to the composite series I created.

Since the post-1975 series imputes costs rather than calculating them directly, it is especially useful to compare them with data published by Greenwich Associates, a firm that has been polling institutional investors on their average commission costs since 1976. The time series of their survey results is plotted in green triangles alongside my own estimates on Figure 6. The two series are

[7] Remarks by A. A. Sommer Jr. in a talk titled "The New Breath of Competition" delivered at the Seminar on the Analysis of Security Prices, University of Chicago, May 15, 1975.

[8] SEC Staff Report on the Securities Industry in 1978

highly similar, except in the first few years of the sample where the commissions paid by institutions are even lower than the computed average. This is consistent with historical reports that the trading commissions charged to individuals did not drop immediately in response to the deregulation until the advent of discount stock brokers around 1980.

Looking at the data prior to 1975, I plot the evolution of the cost of trading a $30 stock using the orange squares. Historical patterns in share prices and trading volume cause the higher frequency variation in my composite series, making it useful to compare against a series where the nominal share price is held constant. Any changes can then be attributed to the imposed cost schedule and not to endogenous investor behavior. Focusing on the cost of trading a $30 stock from 1928 to 1973, we see the round trip cost more than tripled, from 1.07% to 3.46% of the notional value. Including the additional 1.7% for paying the typical $1/4 cost from the bid-ask spread, the total cost of buying and selling exceeded 5% in 1975. It is important to note the economic importance of this magnitude. To put this in perspective, the average stock response to an earnings announcement is in the range of 4%[9], so even if it were possible to know earnings announcements with certainty, you would typically not be able to recover the cost of transacting. The costs were so high that only large misvaluations could merit attention. A speculator would favor low frequency information, with the hope that transaction costs might be amortized over a long horizon. Furthermore, any dynamic trading strategy, such as a portfolio rebalancing rule or a derivative replication, would be incredibly costly.

3.2 Time series analysis

We can expect the constructed time series of transaction costs to be negatively correlated with trading volume, a relationship that should hold true in nearly any economic model. If the proposed efficiency explanation for capital market growth plays a significant role, transaction costs should also be negatively related to capital market spending. In particular, this increase should correspond to active investment management and not just an increase in the operational costs associated with higher trading volume. Lastly, the prediction of more informed speculation also suggests that employees with higher skill and compensation enter the sector in response to a cheaper cost of transacting.

[9] See, for example, Francis, Schipper and Vincent (2002).

The series measuring the cost of capital markets continues to be the value added measure of capital market industries relative to private GDP with annual data from 1927 to 2010. The series measuring capital markets compensation relative to average US private compensation was also previously described and plotted in Figure 1. I measure equity turnover by collecting all available CRSP data on stock volume and shares outstanding for common equity of US firms. Additional details behind the data sources and data construction can be found in the online data appendix.

Summary statistics and simple regression analysis

The summary statistics for these four time series are presented in Table 2. We can see that the transaction cost, measured in basis points (hundredths of one percent), averages 71 basis points over the full sample. The series ranges significantly from more than 150 bps near its peak to just a few basis points in recent years. The fraction of GDP devoted to capital markets averages about 79 basis points over this time series, averaging about 30 basis points before 1975 and increasing to about 200 basis points in recent years. The compensation for capital market employees has an average that is approximately twice the US private sector average over the full sample, increasing to almost 4 times average compensation in recent years. Equity turnover is about 56% a year on average, suggesting an average holding period of approximately two years. While turnover was very high in the late 1920s, it was consistently low for most of the 20th century and then rises again in the recent past, with a current horizon of just a few months.

The correlations of the four series are displayed in the bottom panel of Table 2. As predicted, transaction costs have a strong negative relationship with the size of capital market spending and the volume of trade. While supporting the idea of a contemporaneous relationship, the slow-moving nature of all four time series might cast doubt on the statistical significance.

We can see this more precisely in the simple regressions shown in Table 3, where the GDP share of capital market (*capmkt*), the relative compensation ratio for capital markets (*comp*) and the estimated US equity market turnover (*turnover*) are each regressed on the transaction cost series (*tcost*). As an illustration of the strength of this predictive relationship, Figure 8 plots the growth in the cost of capital markets (shown previously in Figure 1) against the predicted value from the regression. While there is certainly some unexplained variation, the visual fit is striking. Note that each of these series is highly persistent, as is observed in their plots, so it comes as no surprise that an augmented Dickey-Fuler test does not reject the possibility of a unit root. This degree of persistence would discount the significance of their observed correlations.

Regression of first differences

To make a stronger case for this relationship and establish causality (in the Granger sense that past transaction costs forecast growth in capital market activity), we can consider how the changes in one series affects the other by taking first differences. With the high degree of persistence in the raw time series, they may be susceptible to the type of spurious regression results that occur with unit roots. The first differences could then reveal if the time series are truly related, and if so, if one tends to forecast the other. Table 3 reports the results for regressions forecasting annual changes in capital market spending, the capital market compensation ratio, and trading volume as each is regressed on annual changes in transaction costs with up to 4 lags.

The predicted negative relationship remains. Interestingly, changes in transaction costs lead changes in the other series by approximately 2 to 3 years. For example, in the first regression of capital market spending on lagged changes in transaction costs we see negative coefficients for every lag with the second lag being of the strongest magnitude. We can interpret this coefficient as suggesting a one basis point decrease in the cost of transactions predicts that capital markets will consume a 13 basis point higher share of private GDP two years in the future. The same one basis point decrease in the cost of transacting would predict the average compensation of capital markets professionals in three years to rise by an additional 0.18 times the compensation of the average US employee. Looking at the effect on trading volume, this one basis point decrease in transaction costs would suggest trading volume to be 9% higher in three years' time.

This is actually what we might predict if innovations to transaction costs are unexpected. In the context of the proposed model, investors commit to their type ex ante, so we would expect the delayed response to correspond to the time it takes to acquire the talent and research necessary to launch new dynamic strategies.

The statistical relationship seems compelling, although any claims about the importance of the efficiency mechanism are certainly open to critiques of omitted variable bias. A number of important regulatory and technological changes happened during the 1970's. The coincident growth in capital markets and decline in transaction costs could be coincidence, although it would be difficult to explain the strong predictive power of the transaction cost changes exhibited in Table 3. To strengthen the identification of the true mechanism causing financial growth, we can look at the cross-section of firms and focus on specific predictions around the events of May 1975.

4 Market Activity and Asset Prices in the Cross Section

Moving from broad statements about financial activity to the activity we observe for individual firms provides a more refined measure of how much of the growth in active investing can be explained by transaction efficiency. The model presented in section 2 had specific predictions regarding trading activity and the information content of asset prices. As trading efficiency increases we expect to see more trading volume and more informative asset prices. There should also be a differentially large impact on the shorter investment horizons relative to longer horizons. Observing cross-sectional variation in the prices and trading activity of individual firms over the past few decades will generate micro-level support to add to the macro-level time series evidence presented in the previous section.

For increased confidence that we are isolating a key driving mechanism behind the growth of active investing, we can use the events of May 1975 as Rule 19-b came in force. First, we expect that the subsequent drop in transaction costs associated with competitive brokerage commissions should lead to a subsequent increase in the trading and information content of US equities. Following a key prediction of the model, we should expect this to be stronger for shorter horizons. Then, to better identify the efficiency channel, we can use specific features of how the fixed commission schedule affected the cross-section of firms until May 1975 to measure differential effects. This additional level of control helps rule out competing explanations that might have occurred on or around 1975.

4.1 Connecting the panel data with the stylized model

In the stylized model of section 2, the information content of long-horizon prices can be measured through the regression coefficient from projecting the risky investment outcome ($X - \mathrm{E}[X]$) on to the change in the long-horizon price ($R_L = P_0 - P_1$), defining

$$\beta_L = \frac{\mathrm{Cov}[X, R_L]}{\mathrm{Var}[R_L]} = \frac{\beta_{\theta,1}\sigma_\theta^2}{\mathrm{Var}[R_L]}.$$

Intuitively, the information content of long-horizon prices is positively related to the quantity of long-horizon active investors.[10]

The information content of short-horizon prices can be similarly expressed by ($R_S = P_1 - P_2$)

$$\beta_S = \frac{\text{Cov}[X, R_S]}{\text{Var}[R_S]} = \frac{\beta_{\theta,2}\sigma_\theta^2}{\text{Var}[R_S]}.$$

which increases with the sum of the long-horizon and the short-horizon active investors.

We can construct an analogous measure with empirical data on stock prices and earnings. I define the "long horizon" as the period stretching from two years prior to a firm's earnings announcement to 7 months prior to the earnings announcement, the "short horizon" spanning 7 months prior to the earnings announcement to one month prior to the earnings announcement, and the "announcement period" spans from one month before to two months after the announcement. The risky investment outcome will be defined as the scaled change in a firm's quarterly earnings (Δx_t).

This motivates a corresponding empirical regression of the firm's uncertain payout on the returns over each horizon,

$$\Delta x_t = \beta_0 + \beta_L \times r_L + \beta_S \times r_S + \beta_A \times r_A \tag{18}$$

Each of the returns will be measured as the change in log-price, so if time t is measured in months relative to the earnings announcement,

$$\begin{aligned} r_L &= \ln(P_{t-7}) - \ln(P_{t-24}) \\ r_S &= \ln(P_{t-1}) - \ln(P_{t-3})) \\ r_A &= \ln(P_{t+1}) - \ln(P_{t-1})). \end{aligned}$$

Similarly, the risky payout will be measured as a log return scaled by the price observed prior to all the returns. If EPS_t corresponds to the earnings-per-share reported on the announcement date,

[10] Formally, this can be stated as $\frac{\partial \text{Cov}[X,R_L]}{\partial \lambda_L} > 0$, and also, $\frac{\beta_L}{\partial \lambda_L} > 0$ given $\text{Var}[R_L] > \beta_{\theta,1}\sigma_\theta^2$.

the risky payout in the panel regressions specified by (18) will be defined as

$$\Delta x_t = \ln\left(1 + \frac{EPS_t - EPS_{t-3}}{P_{t-24}}\right).$$

4.2 Description of panel data

For each year from 1960 to 2012, I construct a universe of firms by selecting the 1000 largest firms by market capitalization, as measured by their CRSP-reported market cap on December 31st of the prior year. For this set of firms, I collect historical weekly total returns, nominal share prices, trading volume, and shares outstanding. Using the linked CRSP-Compustat data, I collect a panel of their reported earnings per share and the date of the earnings announcement.

The announcements dates are not always available, particularly early in the sample, so I create an additional supplemental series of earnings announcement data where I use historical announcement patterns to estimate the date when not available. This has the advantage of increasing the sample size, and the methodology for estimating historical announcement dates appears to be very accurate when checked against firms for which the actual dates are known. Since the announcement return period is defined to begin one month prior to the reported announcement, any imprecision should have little effect on the results of the subsequent panel regressions.

Table 4 reports the summary statistics for the variables considered in the panel data regression. The earnings news measure (Δx_t) for these large firms over the 45 year sample averages approximately zero with a standard deviation of approximately 2%. The market price for the firms in the sample appears surprisingly high, at about \$104, but this is actually an artifact of Berkshire-Hathaway's inordinately large nominal share price. The median share price is \$32 with a standard deviation of \$24. Dividing the trading volume recorded in CRSP for each quarter by the shares outstanding, I obtain firm-level annualized turnover rates for each firm-quarter in the panel. Over the full sample, annualized turnover averages 2.36, with a wide degree of variation across firms. The return variables, r_L, r_S and r_A, each correspond to a different horizon length, so the magnitudes of their average returns and standard deviations are not directly comparable.

The lower panel of Table 4 reports the same summary statistics for the sub-sample corresponding to the five years before May of 1975, the two years of observations that overlap with May 1975, and five years afterward. This subsample, and ones like it, will be used in the panel regressions where the data window tightens around the events around the implementation of Rule 19-b.

4.3 Rolling panel regression

To generate a graphical measure of the changing information content of prices over time, we can perform a rolling panel regression. I hold the window length constant at two years and then estimate the panel regression corresponding to equation (18) with firm fixed effects. Figure 9 displays the rolling coefficient estimates as a scatterplot in the upper axis, where each estimated long horizon coefficient, β_L, corresponds to a white circle and each estimated short-horizon coefficient, β_S, correspond to a shaded circle. The lower axis reports the estimated root mean square error (RMSE) and the R-squared coefficient of each regression.

The rising pattern in the information content of asset prices is clearly visible. While the magnitude of these betas are roughly similar in the first 10 years of the sample, the predictive power of the short-horizon prices increases much more rapidly than the long-horizon prices. In a more careful subsequent regression estimating the trend in information content over time, I show the increase in the long horizon coefficient, while positive, to be statistically difficult to distinguish from a hypothesis of no change.

This is consistent with the results of Bai et al. (2012). They look at the information content of prices at one to three years prior to earnings releases. This is what my results would consider long-horizon information, and I find no compelling evidence that this information has improved over time.

On the other hand, asset prices less than one year prior to earnings announcements show a consistent increase in information content. Previewing my focus on the events of May 1975, this figure already gives a strong visual indication that the strongest increases in information content correspond to this change as active investing increased dramatically.

While this rolling analysis is instructive, the underlying investment setting may not be fully comparable as the sample rolls across time. The information gathering problem may be different from one decade to the next, and there may be significant changes in the price-to-earnings relationship that would affect the magnitude of the coefficients.

With that in mind, it is interesting to look at the bottom axis of Figure 9 and note how both the explained variation (R^2) and the unexplained variation (RMSE) are increasing in the late 1970's and, to a lesser extent, over the full historical sample. This suggests that the raw difficulty of forecasting earnings increased, but so did the fraction of variation that prices could explain.

4.4 Panel regression with trend

To directly estimate the pattern of change in the information contained in asset prices over the full sample, I run a full panel regression, interacting the return variables with the time trend. The variable, *trend* is measured in years, and the coefficient on $r_L \times trend$ can be interpreted as the annual change in the regression coefficient measuring long-horizon information content. Corresponding interaction terms are used for the short-horizon and announcement return.

Table 5 reports the results of the base panel regressions suggested in equation (18) as well as a version with these time trend interactions. The reported standard errors are estimated using industry clustering, where I use the two digit SIC code as the definition for industry throughout.

The regression reported in the first column of Table 5 reports the results of the base regression using firm fixed effects, considering variation within firms. The second regression specification uses industry and quarter fixed effects to isolate the impact of variation among similar firms in the same time period. The results of each specification are very similar. The strong statistical significance of these regression coefficients should not be too surprising; changes in asset prices correspond to present and future changes in earnings. On the other hand, the coefficient on the long-horizon return is not particularly strong in the first specification with firm fixed effects, and disappears entirely in the second specification.

The third specification is the primary one of interest. It shows the gradual change in these coefficients over time. The interaction term between the short horizon return and the time trend is statistically significant at the 1% level. In contrast the long horizon return shows little evidence of increasing informativeness over time. Of note, the three-month return around the earnings announcement actually shows a decreasing relationship in predicting the reported earnings. The fact that we observe opposite effects on the short-horizon and announcement returns may indicate a substitution of information being pulled into earlier asset prices.

4.5 The post-1975 effect

Over such a long sample, any number of underlying parameters could be changing. The types of firms today are certainly very different than those of the 1960s. There could very well be differences in the difficulty of predicting their future profitability, there can be differences across industries, and there could be differences in their accounting conventions. To be sure that we are truly measuring changes in asset price information and not these other confounding features, we can focus on the change in transaction efficiency associated with the implementation of Rule 19-b in May of 1975 and tighten the estimation window around this period.

I estimate panel regressions using the same framework as before, but I now interact the returns with a dummy variable, $post75$, that equals one for observations where all corresponding variables are observed after the advent of competitive commissions (i.e. after May of 1977). Interacting with this dummy variables tests for a discontinuity in the parameter estimates when crossing this boundary. This regression is reported in Table 6.

There are four regression specifications in the columns of the table, with each one representing a smaller window around 1975. The first specification estimates the panel regression over the full sample, comparing pre-1975 to post-1975 data using the observations from 1966 to 2010. Both long horizon and short horizon prices show dramatic increases in their information content, with their coefficients increasing by a factor of four. However, only the short horizon variables show statistical significance.

The three successive regression specifications with tighter and tighter sample windows increase the standard errors in the coefficient estimates but decrease the concern that other factors unrelated to efficiency and information are driving this result. Looking at the coefficient estimates, the post-1975 effect on short horizon price information remains roughly equal for each time window considered. The effect on long horizon information is always weaker than short horizon and difficult to distinguish from zero.

4.6 Identification using cross-sectional cost differentials

So far the panel analysis has only used the dimension of time to associate active trading and information with transaction efficiency. The strongest evidence for this channel will come from the differential impact across stocks.

The NYSE fixed commission schedule was always a function of the nominal share price. Assuming the nominal share price is a historical artifact, this creates variation across stocks that is plausibly unrelated to any economic characteristics. The commission schedule was set as a decreasing function of nominal share price, so stocks with lower prices were much more expensive to trade than those with higher share prices. This is illustrated in Figure 7, where the round trip cost from the commission schedule effective in May of 1974 is plotted in red on top of a histogram of the frequency distribution of the stock prices at the time.[11]

There are various ways to exploit this variation. The most simplistic is to use a difference in differences approach. I form three categories: $lowP$ for stocks with a nominal share price less than \$15, $midP$ for stocks whose nominal share price is between \$15 and \$30, and $highP$ for stocks whose nominal share price is above \$30. We can then look at the differential impact across categories before and after 1975.

Table 7 reports the results of this approach, where the coefficients of interest are the magnitudes of the product: $r_L \times lowP \times post75$, $r_L \times midP \times post75$, $r_L \times highP \times post75$, $r_S \times highP \times post75$, and so forth. The prediction we are testing is whether these coefficients are positive (indicating more information post-1975) and monotonically decreasing in nominal price (indicating a differential impact across firms according to the relative change in transaction efficiency). As in the previous table, each regression specification corresponds to tighter windows around 1975.

The results for short-horizon prices are just as predicted. All prices appear more informative, but the impact on securities with the largest change in transaction costs ($lowP$) is an order of magnitude higher than stocks where the change was more moderate. As hoped, the relationship is monotonic across the three categories and roughly consistent as the time window shrinks.

In the first regression specification, which uses the longest window, there is some evidence of an increase in information content of long-horizon prices, and the cross-sectional relationship with respect to nominal share price is monotonically decreasing. However, the statistical significance is low, and result disappears entirely in the specifications with shorter sampling windows.

[11] A surprising fact about stock prices is that the distribution of their nominal price per share has been remarkably consistent over time despite inflation and secular changes in investor and investment characteristics. This has been discussed by Weld, Michaely, Thaler and Benartzi (2009).

5 Implications and Conclusions

The empirical analysis shows great success in explaining the modern growth in the cost of capital markets and in looking at its effect on asset prices. However, looking at the information in asset prices only opens the door to broader questions about the social benefits of these changes.

In the simple model presented here, the benefits of active trading largely come from two sources: the noise shocks and the efficient allocation of capital. However, the improved capital allocation is a broadly shared positive externality, not something the active investors accrue directly. The immediate trading profits come at the expense of a counterparty. To what extent will these noise traders be happy in funding trading profits?

5.1 Social welfare

The bigger normative question everyone wants to answer is: are we spending too much on finance? Taking the empirical results back to the modeling framework, we easily see two important welfare effects. First, investors fight over their slice of the pie, leading to what Stein (1987) terms "welfare-reducing speculation." These expenses are wasteful and would suggest too much spending in financial markets. Second, more informed asset prices increase the size of the pie, but the informed investors capture only a small portion of this benefit. All of us who use public market prices are free-riders, and this positive externality suggests we aren't spending nearly enough on informed speculation.

The welfare-reducing speculation can be clearly seen in the simple model where the supply of the risky investment is perfectly inelastic, as it would be for very short horizons. Using the same model parameters that illustrated the equilibrium in section 2, I add a dotted line to the left panel of Figure 10 to show the social welfare (calculated as average expected utility) in the same plot as the expected utility of the active and passive investors. Since the resources spent on information have no effect on total output, social welfare is maximized with practically no informed trading, a solution clearly less than the competitive equilibrium.

It is this type of intuition that drives the suggestions of Philippon (2010), who suggests we may have too few engineers relative to financiers, or Bolton et al. (2011) who similarly contrasts an overabundance of financiers relative to entrepreneurs.

In contrast, the free-riding effect is illustrated in the case of an elastic investment supply, as we would expect for long horizons. The left panel of Figure 11 shows the equilibrium for the same parameters used in the previously discussed example, except the supply of investment will now respond to more accurate asset prices. As you can see, the socially optimal level of informed investment would allocate nearly half of investors to buy information, but the competitive equilibrium allocates far fewer since the uninformed investors are free riding on the social benefits of more informed asset prices.

This analysis builds on the fundamental insight of Hirshleifer (1971), who contrasts the private and social value of foreknowledge. In the model presented here, all information is foreknowledge, learning about information that will inevitably be public knowledge later.

5.2 Conclusions

In the aftermath of the recent financial crisis, scrutiny of financial institutions has increased. The growth in the resources poured into active investment and the surging compensation levels of financial professionals are used as prima facie evidence that financial markets have become inefficient, with many doubting that more active management leads to more informative asset prices.

In a stylized model, I show that investment research and trading are complements, which causes the quantity of both to increase. Financial markets become more informationally and operationally efficient. Empirically, this explanation is very successful in explaining the growth in resources spent in capital markets. Furthermore, it introduces new evidence on the importance of time horizon. Trading horizons have shortened, and there is a corresponding increase in the short-horizon information contained in asset prices.

Since shorter trading horizons may not be socially optimal, this result could be interpreted as justification for Summers and Summers (1989) claim that a non-zero tax on trading might be welfare enhancing, although this requires more explicit measurement of the benefits that arise from informative markets and the recognition that the actual implementation of a financial transaction tax may be impractical (Campbell and Froot, 1994).

The types of dynamic strategies that become feasible with lower transaction costs not only make short-horizon information more valuable but they can also come closer to dynamically completing markets. It is certainly no accident that equity options became widely available in the late 1970s and early 1980s, precisely when US transaction costs experienced their largest drop. The newfound exposures made possible by dynamical hedging may have attracted investors to trade on new risks (Simsek, 2012).

The cost of capital markets has grown enormously over the past few decades. A portion of this can be attributed to the events of May 1975 that enabled dynamic trading strategies and spurred an increase in active investing. This opened the door to modern capital markets, with information and trades moving at ever shorter horizons.

Appendix

A Deriving investor demand

This section of the appendix derives the demand functions for the model with two trading periods ($T = 2$). For each investor, we use their expectations to maximize the utility of final wealth, as defined in equation (7),

$$w_i = w_0 - k_i + q_0 P_1 + q_{i,1}(P_2 - P_1) + q_{i,2}(X - P_2) - \frac{c}{2}\left((q_{i,1} - q_0)^2 + (q_{i,2} - q_{i,1})^2\right).$$

The first order condition, $\frac{\partial}{\partial q_{i,t}} \mathrm{E}_{i,t}[w_i] = \frac{a}{2}\frac{\partial}{\partial q_{i,t}} \mathrm{Var}_{i,t}[w_i]$, can be used to derive the investor demand functions. In period 2, the only source of uncertainty is X and we get

$$q^*_{i,2} = \frac{\mathrm{E}_{i,2}[X - P_2] + cq_{i,1}}{a\mathrm{Var}_{i,2}[X] + c},$$

which leads to the optimal demand functions presented for each type of investor, as in (5).

Deriving the demand functions for period 1 with multiple horizons requires a fair amount of algebra. Beginning with the expression for expected wealth,

$$\mathrm{E}_{i,1}[w_i] = w_0 - k_i + q_0 P_1 + q_{i,1}\mathrm{E}_{i,1}[P_2 - P_1] + \mathrm{E}_{i,1}[q_{i,2}(X - P_2)] - \frac{c}{2}\left((q_{i,1} - q_0)^2 + \mathrm{E}_{i,1}\left[(q_{i,2} - q_{i,1})^2\right]\right),$$

we can substitute in period 2's demand function

$$\begin{aligned}\mathrm{E}_{i,1}[w_i] &= w_0 + q_0 P_1 + q_{i,1}\mathrm{E}_{i,1}[P_2 - P_1] + \mathrm{E}_{i,1}\left[\frac{\mathrm{E}_{i,2}[X - P_2] + cq_{i,1}}{a\mathrm{Var}_{i,2}[X] + c}(X - P_2)\right] \\ &\quad - \frac{c}{2}\left((q_{i,1} - q_0)^2 + \mathrm{E}_{i,1}\left[\left(\frac{\mathrm{E}_{i,2}[X] - P_2 + cq^*_{i,1}}{a\mathrm{Var}_{i,2}[X] + c} - q_{i,1}\right)^2\right]\right)\end{aligned}$$

$$\begin{aligned}\mathrm{E}_{i,1}[w_i] &= w_0 - k_i + q_0 P_1 + q_{i,1}\mathrm{E}_{i,1}[P_2 - P_1] + \frac{\mathrm{E}_{i,1}[X - P_2]}{a\mathrm{Var}_{i,2}[X] + c} cq_{i,1} + \mathrm{E}_{l,1}\left[\frac{\mathrm{E}_{l,2}[X - P_2](X - P_2)}{a\mathrm{Var}_{i,2}[X] + c}\right] \\ &\quad - \frac{c}{2}(q_{i,1} - q_0)^2 - \frac{c}{2}\mathrm{E}_{l,1}\left[\left(\frac{\mathrm{E}_{i,2}[X - P_2]}{a\mathrm{Var}_{i,2}[X] + c} - \frac{a\mathrm{Var}_{i,2}[X]}{a\mathrm{Var}_{i,2}[X] + c} q_{i,1}\right)^2\right]\end{aligned}$$

$$
\begin{aligned}
\mathrm{E}_{i,1}\left[w_i\right] &= w_0 - k_i + q_0 P_1 + q_{i,1}\mathrm{E}_{i,1}\left[P_2 - P_1\right] + \frac{\mathrm{E}_{i,1}\left[X - P_2\right]}{a\mathrm{Var}_{i,2}\left[X\right] + c} c q_{i,1} \\
&\quad + \mathrm{E}_{l,1}\left[\frac{\left(\mathrm{E}_{l,2}\left[X\right] - P_2\right)\left(X - P_2\right)}{a\mathrm{Var}_{i,2}\left[X\right] + c}\right] - \frac{c}{2}\left(q_{i,1} - q_0\right)^2 - \frac{c}{2}\left(\frac{a\mathrm{Var}_{i,2}\left[X\right]}{a\mathrm{Var}_{i,2}\left[X\right] + c}\right)^2 q_{i,1}^2 \\
&\quad - \frac{c}{2}\mathrm{E}_{l,1}\left[\left(\frac{\mathrm{E}_{i,2}\left[X - P_2\right]}{a\mathrm{Var}_{i,2}\left[X\right] + c}\right)^2\right] + c\frac{\mathrm{E}_{i,1}\left[X - P_2\right] a\mathrm{Var}_{i,2}\left[X\right]}{\left(a\mathrm{Var}_{i,2}\left[X\right] + c\right)^2} q_{i,1}
\end{aligned}
$$

with first derivative

$$
\begin{aligned}
\frac{\partial}{\partial q_{i,1}}\mathrm{E}\left[w_i\right] &= \mathrm{E}_{i,1}\left[P_2 - P_1\right] + \frac{c\mathrm{E}_{i,1}\left[X - P_2\right]}{a\mathrm{Var}_{i,2}\left[X\right] + c} - c\left(q_{i,1} - q_0\right) \\
&\quad - c\left(\frac{a\mathrm{Var}_{i,2}\left[X\right]}{a\mathrm{Var}_{i,2}\left[X\right] + c}\right)^2 q_{i,1} + c\frac{a\mathrm{Var}_{i,2}\left[X\right]\mathrm{E}_{l,1}\left[X - P_2\right]}{\left(a\mathrm{Var}_{i,2}\left[X\right] + c\right)^2}
\end{aligned}
$$

so the final expression is

$$
\begin{aligned}
\frac{\partial}{\partial q_{i,1}}\mathrm{E}\left[w_i\right] &= \mathrm{E}_{i,1}\left[P_2 - P_1\right] + \left(\frac{c}{a\mathrm{Var}_{i,2}\left[X\right] + c} + c\frac{a\mathrm{Var}_{i,2}\left[X\right]}{\left(a\mathrm{Var}_{i,2}\left[X\right] + c\right)^2}\right)\mathrm{E}_{i,1}\left[X - P_2\right] + c q_0 \\
&\quad - c\left(1 + \left(\frac{a\mathrm{Var}_{i,2}\left[X\right]}{a\mathrm{Var}_{i,2}\left[X\right] + c}\right)^2\right) q_{i,1}
\end{aligned}
$$

The optimal portfolio in period one will be the one that solves the first order condition,

$$
q_{i,1}^* = \frac{\mathrm{E}_{i,1}\left[P_2 - P_1\right] + \left(\frac{c}{a\mathrm{Var}_{i,2}[X]+c} + c\frac{a\mathrm{Var}_{i,2}[X]}{(a\mathrm{Var}_{i,2}[X]+c)^2}\right)\mathrm{E}_{i,1}\left[X - P_2\right] + c q_0}{\frac{a}{2q_{i,1}^*}\mathrm{Var}_{i,1}\left[w_i\right] + \left(1 + \left(\frac{a\mathrm{Var}_{i,2}[X]}{a\mathrm{Var}_{i,2}[X]+c}\right)^2\right) c}.
$$

The expected values for P_2 and X are apparent from the assumed linearity in (8) and (9), so the task at hand is to come up with expressions for $\frac{a}{2q_{i,1}^*}\mathrm{Var}_{i,1}\left[w_3\right]$, where the variance term can be expressed as

$$
\begin{aligned}
\mathrm{Var}_{i,1}\left[w_i\right] &= \mathrm{Var}_{i,1}\left[q_{i,1}P_2 + q_{i,2}\left(X - P_2\right) - \frac{c}{2}\left(q_{i,2} - q_{i,1}\right)^2\right] \\
&= \mathrm{Var}_{i,1}\left[q_{i,1}P_2 + q_{i,2}\left(X - P_2\right) - \frac{c}{2}q_{i,2}^2 + c q_{l,1} q_{i,2}\right] \\
&= \mathrm{Var}_{i,1}\left[q_{i,1}P_2 + q_{i,2}\left(q_{i,2}\left(a\mathrm{Var}_{i,2}\left[X\right] + c\right) + \left(X - \mathrm{E}_{i,2}\left[X\right]\right)\right) - \frac{c}{2}q_{i,2}^2\right] \\
&= \mathrm{Var}_{i,1}\left[q_{i,1}P_2 + q_{i,2}\left(X - \mathrm{E}_{i,2}\left[X\right]\right) + q_{i,2}^2\left(a\mathrm{Var}_{i,2}\left[X\right] + \frac{c}{2}\right)\right]
\end{aligned}
$$

and the remaining calculation requires using the expectations of each investor and calculating the sensitivity with respect to the first period allocation..

A.1 Long-horizon investors in period 1

For long-horizon investors, the uncertain terms will be:

$$P_2 - \mathrm{E}_{L,1}[P_2] = \beta_{\nu_2,2}\nu_2,$$

$$X - \mathrm{E}_{L,1}[X] = X - \mathrm{E}_{l,2}[X] = \varepsilon.$$

The optimal position during the final trading period

$$\begin{aligned} q_{L,2} &= \frac{\mathrm{E}_{L,1}[X-P_2] + cq_{L,1}}{a\sigma_\varepsilon^2 + c} - \frac{\beta_{\nu_2}\nu_2}{a\sigma_\varepsilon^2 + c} \\ &= \mathrm{E}_{L,1}[q_{L,2}] - \frac{\beta_{\nu_2}}{a\sigma_\varepsilon^2 + c}\nu_2. \end{aligned}$$

From this, we can calculate the variance

$$\begin{aligned} \mathrm{Var}_{L,1}[w_L] &= \mathrm{Var}_{L,1}\left[q_{L,1}\beta_{\nu_2}\nu_2 + \left(\mathrm{E}_{L,1}[q_{L,2}] - \frac{\beta_{\nu_2}}{a\sigma_\varepsilon^2 + c}\nu_2\right)\varepsilon + \left(\mathrm{E}_{L,1}[q_{L,2}] - \frac{\beta_{\nu_2}}{a\sigma_\varepsilon^2 + c}\nu_2\right)^2\left(a\sigma_\varepsilon^2 + \frac{c}{2}\right)\right] \\ &= \mathrm{Var}_{L,1}\left[\begin{array}{c} q_{L,1}\beta_{\nu_2}\nu_2 + \mathrm{E}_{L,2}[q_{L,2}]\varepsilon - \frac{\beta_{\nu_2}}{a\sigma_\varepsilon^2+c}\nu_2\varepsilon \\ +\left(\mathrm{E}_{L,1}[q_{L,2}] - \frac{\beta_{\nu_2}}{a\sigma_\varepsilon^2+c}\nu_2\right)^2\left(a\sigma_\varepsilon^2 + \frac{c}{2}\right) \end{array}\right] \\ &= \mathrm{Var}_{L,1}\left[\begin{array}{c} \left(\frac{c}{a\sigma_\varepsilon^2+c}q_{L,1} + \frac{\mathrm{E}_{L,1}[X-P_2]}{a\sigma_\varepsilon^2+c}\right)\varepsilon \\ +\left(q_{L,1}\left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2+c}\right)^2 - 2\mathrm{E}_{L,1}[X-P_2]\frac{a\sigma_\varepsilon^2+\frac{c}{2}}{(a\sigma_\varepsilon^2+c)^2}\right)\beta_{\nu_2}\nu_2 \\ -\frac{\beta_{\nu_2}}{a\sigma_\varepsilon^2+c}\nu_2\varepsilon + \frac{a\sigma_\varepsilon^2+\frac{c}{2}}{(a\sigma_\varepsilon^2+c)^2}\beta_{\nu_2}^2\nu_2^2 \end{array}\right] \end{aligned}$$

and using the normality and independence of ε and ν_2,

$$\begin{aligned} \mathrm{Var}[w_L] &= \mathrm{Var}[a\nu + b\varepsilon + c\nu^2 + d\nu\varepsilon] \\ &= a^2\sigma_\nu^2 + b^2\sigma_\varepsilon^2 + 2c^2\sigma_\nu^4 + d^2\sigma_\nu^2\sigma_\varepsilon^2 \end{aligned}$$

we can write

$$
\begin{aligned}
\mathrm{Var}_{L,1}[w_L] &= \left(\frac{c}{a\sigma_\varepsilon^2+c}q_{i,1}+\frac{\mathrm{E}_{L,1}[X-P_2]}{a\sigma_\varepsilon^2+c}\right)^2\sigma_\varepsilon^2 \\
&\quad+\left(q_{L,1}\left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2+c}\right)^2-\mathrm{E}_{L,1}[X-P_2]\frac{2a\sigma_\varepsilon^2+c}{(a\sigma_\varepsilon^2+c)^2}\right)^2\beta_{\nu_2}^2\sigma_\nu^2 \\
&\quad+\left\{\frac{\beta_{\nu_2}}{a\sigma_\varepsilon^2+c}\right\}^2\sigma_\nu^2\sigma_\varepsilon^2+2\left\{\frac{a\sigma_\varepsilon^2+\frac{c}{2}}{(a\sigma_\varepsilon^2+c)^2}\right\}^2\beta_{\nu_2,2}^4\sigma_2^4.
\end{aligned}
$$

To calculate the demand function, we need to evaluate the first derivative

$$
\begin{aligned}
\frac{\partial}{\partial q_{1,l}}\mathrm{Var}_{i,1}[w_i] &= 2\frac{c}{a\sigma_\varepsilon^2+c}\left(\frac{c}{a\sigma_\varepsilon^2+c}q_{i,1}+\frac{\mathrm{E}_{L,1}[X-P_2]}{a\sigma_\varepsilon^2+c}\right)\sigma_\varepsilon^2 \\
&\quad+2\left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2+c}\right)^2\left(q_{i,1}\left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2+c}\right)^2-\mathrm{E}_{L,1}[X-P_2]\frac{2a\sigma_\varepsilon^2+c}{(a\sigma_\varepsilon^2+c)^2}\right)\beta_{\nu_2}^2\sigma_\nu^2
\end{aligned}
$$

and calculate the term

$$
\begin{aligned}
\frac{a}{2}\frac{\partial\mathrm{Var}_{L,1}[w_L]}{\partial q_{1,l}} &= a\left(\left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2+c}\right)^4\beta_{\nu_2}^2\sigma_\nu^2+\left(\frac{c}{a\sigma_\varepsilon^2+c}\right)^2\sigma_\varepsilon^2\right)q_{i,1} \\
&\quad-a\left(\frac{\left(2a\sigma_\varepsilon^2+c\right)a^2\sigma_\varepsilon^4\beta_{\nu_2}^2\sigma_\nu^2}{(a\sigma_\varepsilon^2+c)^4}-\frac{c\sigma_\varepsilon^2}{(a\sigma_\varepsilon^2+c)^2}\right)\mathrm{E}_{L,2}[X-P_2]
\end{aligned}
$$

The optimal portfolio for the long-term speculator is then

$$
q_{L,1}^*=\frac{\mathrm{E}_{L,1}[P_2-P_1]+\left\{\frac{c}{a\sigma_\varepsilon^2+c}+a\frac{\left(2a\sigma_\varepsilon^2+c\right)\beta_{\nu_2,2}^2\sigma_\nu^2a^2\sigma_\varepsilon^4}{(a\sigma_\varepsilon^2+c)^4}\right\}\mathrm{E}_{L,1}[X-P_2]+cq_0}{a\left\{\left(\frac{c}{a\sigma_\varepsilon^2+c}\right)^2\sigma_\varepsilon^2+\left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2+c}\right)^4\beta_{\nu_2,2}^2\sigma_\nu^2\right\}+c\left(1-\left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2+c}\right)^2\right)}
$$

which can be written as in equation (17)

$$
\begin{aligned}
q_{L,1}^* &= \frac{\mathrm{E}_{L,1}[P_2-P_1]+\Gamma_l\mathrm{E}_{L,1}[X-P_2]+cq_0}{a\Omega+c\left(1-\left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2+c}\right)^2\right)} \\
&= \frac{(1-\Gamma)\,\mathrm{E}_{L,1}[P_2-P_1]+\Gamma\mathrm{E}_{L,1}[X-P_1]+cq_0}{a\Omega+c\left(1-\left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2+c}\right)^2\right)}
\end{aligned}
$$

The variance term, Ω is a linear combination of the uncertainty in next period's price (σ_ν^2) and uncertainty in the final payout (σ_ε^2)

$$\Omega = \underbrace{\left(\frac{c}{a\sigma_\varepsilon^2 + c}\right)^2 \sigma_\varepsilon^2}_{\text{variance of } X} + \underbrace{\left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2 + c}\right)^4 \beta_{\nu_2,2}^2 \sigma_\nu^2}_{\text{variance in } P_2}.$$

The sensitivity to next period's expected return is

$$\Gamma = \underbrace{\frac{c}{a\sigma_\varepsilon^2 + c}}_{\text{return next period}} + \underbrace{a\frac{\left(2a\sigma_\varepsilon^2 + c\right) a^2 \sigma_\varepsilon^4}{\left(a\sigma_\varepsilon^2 + c\right)^4} \beta_{\nu_2,2}^2 \sigma_\nu^2}_{\text{prefer to avoid uncertain } \nu_2}.$$

The weight Γ that the investor tilts toward the long-horizon return will always be positive, and its magnitude will increase with transaction costs. The relationship with transaction costs comes from the investor recognizing positions taken now will persist later. Additionally, there is some uncertainty in the price next period, so investors have an incentive to lock in P_1 now rather than pay an uncertain P_2.

A.2 Short-horizon investors in period 1

For the short-run investors, the uncertain terms will be

$$P_2 - \mathrm{E}_{N,1}\left[P_2\right] = \beta_4 e_S + \beta_{\nu_2} \nu_2$$

and

$$X - \mathrm{E}_{s,1}\left[X\right] = e_S + \varepsilon,$$

where

$$e_S = \left(\theta_2 - \mathrm{E}_{S,1}\left[\theta_2\right]\right).$$

The optimal portfolio in the final trading period can then be expressed as

$$
\begin{aligned}
q_{S,2} &= \frac{\mathrm{E}_{S,2}\left[X-P_2\right]+cq_{S,1}}{a\sigma_\varepsilon^2+c} \\
&= \frac{\mathrm{E}_{S,1}\left[X-P_2\right]+cq_{S,1}}{a\sigma_\varepsilon^2+c}+\frac{\mathrm{E}_{S,2}\left[X-P_2\right]-\mathrm{E}_{S,1}\left[X-P_2\right]}{a\sigma_\varepsilon^2+c} \\
&= \mathrm{E}_{S,1}\left[q_{S,2}\right]+\frac{e_S\left(1-\beta_4\right)}{a\sigma_\varepsilon^2+c}-\frac{\beta_{\nu_2}\nu_2}{a\sigma_\varepsilon^2+c}.
\end{aligned}
$$

So we can calculate the variance as

$$
\begin{aligned}
\mathrm{Var}_{S,1}\left[w_S\right] &= \mathrm{Var}_{S,1}\left[q_{S,1}P_2+q_{S,2}\left(X-\mathrm{E}_{S,2}\left[X\right]\right)+q_{S,2}^2\left(a\sigma_\varepsilon^2+\frac{c}{2}\right)\right] \\
&= \mathrm{Var}_{S,1}\left[\begin{array}{c} q_{S,1}\left(\beta_4 e_S+\beta_{\nu_2}\nu_2\right)+\left(\mathrm{E}_{S,1}\left[q_{S,2}\right]+\frac{e_S(1-\beta_4)}{a\sigma_\varepsilon^2+c}-\frac{\beta_{\nu_2}\nu_2}{a\sigma_\varepsilon^2+c}\right)\varepsilon \\ +\left(\mathrm{E}_{S,1}\left[q_{S,2}\right]+\frac{e_S(1-\beta_4)}{a\sigma_\varepsilon^2+c}-\frac{\beta_{\nu_2}\nu_2}{a\sigma_\varepsilon^2+c}\right)^2\left(a\sigma_\varepsilon^2+\frac{c}{2}\right) \end{array}\right]
\end{aligned}
$$

$$
\mathrm{Var}_{S,1}\left[w_S\right]=\mathrm{Var}_{S,1}\left[\begin{array}{c} \left(q_{S,1}-2\frac{(1-\beta_4)\mathrm{E}_{S,1}\left[q_{S,2}\right]\left(a\sigma_\varepsilon^2+\frac{c}{2}\right)}{a\sigma_\varepsilon^2+c}\right)\beta_4 e_S \\ +\left(q_{S,1}-2\frac{\mathrm{E}_{S,1}\left[q_{S,2}\right]\left(a\sigma_\varepsilon^2+\frac{c}{2}\right)}{a\sigma_\varepsilon^2+c}\right)\beta_{\nu_2}\nu_2+\mathrm{E}_{S,1}\left[q_{S,2}\right]\varepsilon \\ +\left(\frac{e_S(1-\beta_4)}{a\sigma_\varepsilon^2+c}-\frac{\beta_{\nu_2}\nu_2}{a\sigma_\varepsilon^2+c}\right)^2\left(a\sigma_\varepsilon^2+\frac{c}{2}\right)+\left(\frac{e_S(1-\beta_4)}{a\sigma_\varepsilon^2+c}-\frac{\beta_{\nu_2}\nu_2}{a\sigma_\varepsilon^2+c}\right)\varepsilon \end{array}\right]
$$

So the variance is

$$
\begin{aligned}
\mathrm{Var}_{S,1}\left[w_S\right] &= \left(q_{S,1}-2\frac{\left(1-\beta_4\right)\mathrm{E}_{S,1}\left[q_{S,2}\right]\left(a\sigma_\varepsilon^2+\frac{c}{2}\right)}{a\sigma_\varepsilon^2+c}\right)^2\beta_4^2\sigma_{S,1}^2 \\
&+\left(q_{S,1}-2\frac{\mathrm{E}_{S,1}\left[q_{S,2}\right]\left(a\sigma_\varepsilon^2+\frac{c}{2}\right)}{a\sigma_\varepsilon^2+c}\right)^2\beta_{\nu_2}^2\sigma_\nu^2 \\
&+\left(\mathrm{E}_{S,1}\left[q_{S,2}\right]\right)^2\sigma_\varepsilon^2 \\
&+\mathrm{Var}_{S,1}\left[\left(\frac{e_S\left(1-\beta_4\right)}{a\sigma_\varepsilon^2+c}-\frac{\beta_{\nu_2}\nu_2}{a\sigma_\varepsilon^2+c}\right)^2\left(a\sigma_\varepsilon^2+\frac{c}{2}\right)+\left(\frac{e_S\left(1-\beta_4\right)}{a\sigma_\varepsilon^2+c}-\frac{\beta_{\nu_2}\nu_2}{a\sigma_\varepsilon^2+c}\right)\varepsilon\right]
\end{aligned}
$$

and we can substitute in $\mathrm{E}_{S,1}\left[q_{S,2}\right]=\frac{\mathrm{E}_{S,1}[X-P_2]+cq_{S,1}}{a\sigma_\varepsilon^2+c}$ get

$$\begin{aligned}
\mathrm{Var}_{S,1}\left[w_S\right] &= \left(q_{S,1}\left(\frac{a^2\sigma_\varepsilon^4-\beta_4\left(2ac\sigma_\varepsilon^2+c^2\right)}{\left(a\sigma_\varepsilon^2+c\right)^2}\right)-\frac{\left(1-\beta_4\right)\left(2a\sigma_\varepsilon^2+c\right)}{\left(a\sigma_\varepsilon^2+c\right)^2}\mathrm{E}_{S,1}\left[X-P_2\right]\right)^2\beta_4^2\sigma_{S,1}^2 \\
&+\left(q_{S,1}\frac{a^2\sigma_\varepsilon^4}{\left(a\sigma_\varepsilon^2+c\right)^2}-\frac{2a\sigma_\varepsilon^2+c}{\left(a\sigma_\varepsilon^2+c\right)^2}\mathrm{E}_{S,1}\left[X-P_2\right]\right)^2\beta_{\nu_2}^2\sigma_\nu^2 \\
&+\left(q_{S,1}\frac{c}{a\sigma_\varepsilon^2+c}+\frac{\mathrm{E}_{S,1}\left[X-P_2\right]}{a\sigma_\varepsilon^2+c}\right)^2\sigma_\varepsilon^2 \\
&+\mathrm{Var}_{S,1}\left[\left(\frac{e_S\left(1-\beta_4\right)}{a\sigma_\varepsilon^2+c}-\frac{\beta_{\nu_2}\nu_2}{a\sigma_\varepsilon^2+c}\right)^2\left(a\sigma_\varepsilon^2+\frac{c}{2}\right)+\left(\frac{e_S\left(1-\beta_4\right)}{a\sigma_\varepsilon^2+c}-\frac{\beta_{\nu_2}\nu_2}{a\sigma_\varepsilon^2+c}\right)\varepsilon\right]
\end{aligned}$$

with first derivative

$$\begin{aligned}
\frac{\partial\mathrm{Var}_{S,1}\left[w_S\right]}{\partial q_{S,1}} &= 2q_{S,1}\left\{\left(\frac{a^2\sigma_\varepsilon^4}{\left(a\sigma_\varepsilon^2+c\right)^2}\right)^2\beta_{\nu_2}^2\sigma_\nu^2+\left(\frac{c}{a\sigma_\varepsilon^2+c}\right)^2\sigma_\varepsilon^2\right\} \\
&+2q_{S,1}\left\{\left(\frac{a^2\sigma_\varepsilon^4-\beta_4\left(2ac\sigma_\varepsilon^2+c^2\right)}{\left(a\sigma_\varepsilon^2+c\right)^2}\right)^2\beta_4^2\sigma_{S,1}^2\right\} \\
&-2\mathrm{E}_{S,1}\left[X-P_2\right]\left(+\frac{a^2\sigma_\varepsilon^4\left(2a\sigma_\varepsilon^2+c\right)}{\left(a\sigma_\varepsilon^2+c\right)^4}\beta_{\nu_2}^2\sigma_\nu^2-\frac{c\sigma_\varepsilon^2}{\left(a\sigma_\varepsilon^2+c\right)^2}\right) \\
&-2\mathrm{E}_{S,1}\left[X-P_2\right]\left(\frac{\left(a^2\sigma_\varepsilon^4-\beta_4\left(2ac\sigma_\varepsilon^2+c^2\right)\right)\left(1-\beta_4\right)\left(2a\sigma_\varepsilon^2+c\right)}{\left(a\sigma_\varepsilon^2+c\right)^4}\beta_4^2\sigma_{S,1}^2\right)
\end{aligned}$$

The optimal portfolio is then

$$q_{S,1}^*=\frac{\mathrm{E}_{S,1}\left[P_2-P_1\right]+\left(\frac{c}{a\sigma_\varepsilon^2+c}+\frac{\left(2a\sigma_\varepsilon^2+c\right)\left(a^2\sigma_\varepsilon^4\beta_{\nu_2}^2\sigma_\nu^2+\left(a^2\sigma_\varepsilon^4-\beta_4\left(2ac\sigma_\varepsilon^2+c^2\right)\right)\left(1-\beta_4\right)\beta_4^2\sigma_{S,1}^2\right)}{\left(a\sigma_\varepsilon^2+c\right)^4}\right)\mathrm{E}_{S,1}\left[X-P_2\right]+cq_0}{a\left(\left(\frac{c}{a\sigma_\varepsilon^2+c}\right)^2\sigma_\varepsilon^2+\frac{\left(a^2\sigma_\varepsilon^4-\beta_4\left(2ac\sigma_\varepsilon^2+c^2\right)\right)^2\beta_4^2\sigma_{S,1}^2+a^4\sigma_\varepsilon^8\beta_{\nu_2}^2\sigma_\nu^2}{\left(a\sigma_\varepsilon^2+c\right)^4}\right)+\left(1+\left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2+c}\right)^2\right)c}$$

which can be expressed in a form analogous to the long-run demand function in equation (17) by naming the short-horizon parameters, Γ_s and Ω_s,

$$q_{S,1}^*=\frac{\left(1-\Gamma_S\right)\mathrm{E}_{S,1}\left[P_2-P_1\right]+\Gamma_S\mathrm{E}_{S,1}\left[X-P_1\right]+cq_0}{a\Omega_S+c\left(1+\left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2+c}\right)^2\right)}. \tag{19}$$

The intuition and form are nearly identical, with the short-horizon investors tilting slightly more toward the long-run return, $\mathrm{E}_{S,1}[X-P_2]$, due to their uncertainty about θ_2,

$$\Gamma_S = \Gamma + \underbrace{a\frac{(1-\beta_4)\left(a^2\sigma_\varepsilon^4 - \beta_4\left(2ac\sigma_\varepsilon^2 + c^2\right)\right)}{\left(a\sigma_\varepsilon^2 + c\right)^4}\beta_4^2\sigma_{e_1}^2}_{\text{prefer to avoid uncertain } e_1}. \tag{20}$$

A.3 Uninformed investors in period 1

The uninformed investors have the highest degree of uncertainty. In period 1, this is summarized by the uncertain terms:

$$X - \mathrm{E}_{N,2}[X] = e_1 + e_2 + \varepsilon$$

where the errors in expectations in the final period are expressed as

$$\begin{aligned} e_1 &= (\theta_1 - \mathrm{E}_{N,2}[\theta_1]) \\ e_2 &= (\theta_2 - \mathrm{E}_{N,2}[\theta_2]). \end{aligned}$$

The additional, orthogonal error in the first period expectation is

$$\begin{aligned} \Delta e_1 &= (\theta_1 - \mathrm{E}_{N,1}[\theta_1]) - (\theta_1 - \mathrm{E}_{N,2}[\theta_1]) \\ \Delta e_2 &= (\theta_2 - \mathrm{E}_{N,1}[\theta_2]) - (\theta_2 - \mathrm{E}_{N,2}[\theta_2]) \end{aligned}$$

so that

$$P_2 - \mathrm{E}_{N,1}[P_2] = \beta_3(e_1 + \Delta e_1) + \beta_4(e_2 + \Delta e_2) + \beta_{\nu_2}\nu_2$$

and

$$X - \mathrm{E}_{N,1}[X] = (e_1 + \Delta e_1) + (e_2 + \Delta e_2) + \varepsilon.$$

The optimal portfolio in the final trading period can then be expressed as

$$\begin{aligned}
q_{N,2} &= \frac{\mathrm{E}_{N,2}\left[X-P_2\right]+cq_{N,1}}{a\mathrm{Var}_{N,2}\left[X\right]+c} \\
&= \frac{\mathrm{E}_{N,1}\left[X-P_2\right]+cq_{N,1}}{a\mathrm{Var}_{N,2}\left[X\right]+c}+\frac{\mathrm{E}_{N,2}\left[X-P_2\right]-\mathrm{E}_{N,1}\left[X-P_2\right]}{a\mathrm{Var}_{N,2}\left[X\right]+c} \\
&= \frac{\mathrm{E}_{N,1}\left[X-P_2\right]+cq_{N,1}}{a\mathrm{Var}_{N,2}\left[X\right]+c}+\frac{\mathrm{E}_{N,2}\left[X\right]-\mathrm{E}_{N,1}\left[X\right]-P_2-\mathrm{E}_{N,1}\left[P_2\right]}{a\mathrm{Var}_{N,2}\left[X\right]+c} \\
&= \mathrm{E}_{N,1}\left[q_{N,2}\right]+\frac{\Delta e_1+\Delta e_2-\beta_3\left(e_1+\Delta e_1\right)-\beta_4\left(e_2+\Delta e_2\right)-\beta_{\nu_2}\nu_2}{a\mathrm{Var}_{N,2}\left[X\right]+c} \\
&= \mathrm{E}_{N,1}\left[q_{N,2}\right]+\frac{\Delta e_1\left(1-\beta_3\right)+\Delta e_2\left(1-\beta_4\right)-\beta_3 e_1-\beta_4 e_2-\beta_{\nu_2}\nu_2}{a\mathrm{Var}_{N,2}\left[X\right]+c}.
\end{aligned}$$

The uncertainty from the perspective of the investors who acquire no information will be

$$\begin{aligned}
\mathrm{Var}_{N,1}\left[w_N\right] &= \mathrm{Var}_{N,1}\left[q_{N,1}P_2+q_{N,2}\left(X-\mathrm{E}_{N,2}\left[X\right]\right)+q_{N,2}^2\left(a\mathrm{Var}_{N,2}\left[X\right]+\frac{c}{2}\right)\right] \\
&= \mathrm{Var}_{N,1}\left[\begin{array}{c}
q_{N,1}\left(\beta_3\left(e_1+\Delta e_1\right)+\beta_4\left(e_2+\Delta e_2\right)+\beta_{\nu_2}\nu_2\right) \\
+\left(\mathrm{E}_{N,1}\left[q_{N,2}\right]+\frac{\Delta e_1(1-\beta_3)+\Delta e_2(1-\beta_4)-\beta_3 e_1-\beta_4 e_2-\beta_{\nu_2}\nu_2}{a\mathrm{Var}_{N,2}[X]+c}\right)\left(e_1+e_2+\varepsilon\right) \\
+\left(\mathrm{E}_{N,1}\left[q_{N,2}\right]+\frac{\Delta e_1(1-\beta_3)+\Delta e_2(1-\beta_4)-\beta_3 e_1-\beta_4 e_2-\beta_{\nu_2}\nu_2}{a\mathrm{Var}_{N,2}[X]+c}\right)^2\left(a\mathrm{Var}_{N,2}\left[X\right]+\frac{c}{2}\right)
\end{array}\right]
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}_{N,1}\left[w_N\right] = &\left\{q_{N,1}\beta_3+\mathrm{E}_{N,1}\left[q_{N,2}\right]-2\beta_3\mathrm{E}_{N,1}\left[q_{N,2}\right]\frac{a\mathrm{Var}_{N,2}\left[X\right]+\frac{c}{2}}{a\mathrm{Var}_{N,2}\left[X\right]+c}\right\}^2\sigma_{e_1}^2 \\
&+\left\{q_{N,1}\beta_4+\mathrm{E}_{N,1}\left[q_{N,2}\right]-2\beta_4\mathrm{E}_{N,1}\left[q_{N,2}\right]\frac{a\mathrm{Var}_{N,2}\left[X\right]+\frac{c}{2}}{a\mathrm{Var}_{N,2}\left[X\right]+c}\right\}^2\sigma_{e_2}^2 \\
&+\left\{q_{N,1}\beta_3+\left(1-\beta_3\right)2\mathrm{E}_{N,1}\left[q_{N,2}\right]\frac{a\mathrm{Var}_{N,2}\left[X\right]+\frac{c}{2}}{a\mathrm{Var}_{N,2}\left[X\right]+c}\right\}^2\sigma_{\Delta e_1}^2 \\
&+\left\{q_{N,1}\beta_4+\left(1-\beta_4\right)2\mathrm{E}_{N,1}\left[q_{N,2}\right]\frac{a\mathrm{Var}_{N,2}\left[X\right]+\frac{c}{2}}{a\mathrm{Var}_{N,2}\left[X\right]+c}\right\}^2\sigma_{\Delta e_2}^2 \\
&+\left\{q_{N,1}-2\mathrm{E}_{N,1}\left[q_{N,2}\right]\frac{a\mathrm{Var}_{N,2}\left[X\right]+\frac{c}{2}}{a\mathrm{Var}_{N,2}\left[X\right]+c}\right\}^2\beta_{\nu_2}^2\sigma_{\nu_2}^2 \\
&+\left\{\mathrm{E}_{N,1}\left[q_{N,2}\right]\right\}^2\sigma_{\varepsilon}^2 \\
&+\left\{\text{the terms without } q_{N,1}\right\}
\end{aligned}$$

which we can rewrite to focus on $q_{N,1}$ as

$$
\begin{aligned}
\mathrm{Var}_{N,1}[w_N] &= \left\{ \begin{array}{c} q_{N,1}\left(\frac{c}{a\mathrm{Var}_{N,2}[X]+c} + \beta_3\left(\frac{a\mathrm{Var}_{N,2}[X]}{a\mathrm{Var}_{N,2}[X]+c}\right)^2\right) \\ +\mathrm{E}_{N,1}[X-P_2]\left(\frac{a\mathrm{Var}_{N,2}[X]+c-2\beta_3\left(a\mathrm{Var}_{N,2}[X]+\frac{c}{2}\right)}{\left(a\mathrm{Var}_{N,2}[X]+c\right)^2}\right) \end{array} \right\}^2 \sigma_{e_1}^2 \\
&+ \left\{ \begin{array}{c} q_{N,1}\left(\frac{c}{a\mathrm{Var}_{N,2}[X]+c} + \beta_4\left(\frac{a\mathrm{Var}_{N,2}[X]}{a\mathrm{Var}_{N,2}[X]+c}\right)^2\right) \\ +\mathrm{E}_{N,1}[X-P_2]\left(\frac{a\mathrm{Var}_{N,2}[X]+c-2\beta_4\left(a\mathrm{Var}_{N,2}[X]+\frac{c}{2}\right)}{\left(a\mathrm{Var}_{N,2}[X]+c\right)^2}\right) \end{array} \right\}^2 \sigma_{e_2}^2 \\
&+ \left\{ \begin{array}{c} q_{N,1}\left(1-\beta_3\left(\frac{a\mathrm{Var}_{N,2}[X]}{a\mathrm{Var}_{N,2}[X]+c}\right)^2\right) \\ +\mathrm{E}_{N,1}[X-P_2]\left(\frac{(1-\beta_3)\left(2a\mathrm{Var}_{N,2}[X]+c\right)}{\left(a\mathrm{Var}_{N,2}[X]+c\right)^2}\right) \end{array} \right\}^2 \sigma_{\Delta e_1}^2 \\
&+ \left\{ \begin{array}{c} q_{N,1}\left(1-\beta_4\left(\frac{a\mathrm{Var}_{N,2}[X]}{a\mathrm{Var}_{N,2}[X]+c}\right)^2\right) \\ +\mathrm{E}_{N,1}[X-P_2]\left(\frac{(1-\beta_4)\left(2a\mathrm{Var}_{N,2}[X]+c\right)}{\left(a\mathrm{Var}_{N,2}[X]+c\right)^2}\right) \end{array} \right\}^2 \sigma_{\Delta e_2}^2 \\
&+ \left\{ q_{N,1}\left(1-\left(\frac{a\mathrm{Var}_{N,2}[X]}{a\mathrm{Var}_{N,2}[X]+c}\right)^2\right) - \mathrm{E}_{N,1}[X-P_2]\left(\frac{2a\mathrm{Var}_{N,2}[X]+c}{(a\mathrm{Var}_{N,2}[X]+c)^2}\right)\right\}^2 \beta_{\nu_2}^2\sigma_{\nu_2}^2 \\
&+ \left\{ q_{N,1}\frac{c}{a\mathrm{Var}_{N,2}[X]+c} + \frac{\mathrm{E}_{N,1}[X-P_2]}{a\mathrm{Var}_{N,2}[X]+c}\right\}^2 \sigma_\varepsilon^2 \\
&+ \{\text{the terms without } q_{N,1}\}
\end{aligned}
$$

and taking the first derivative yields the comon form

$$
q_{N,1}^* = \frac{(1-\Gamma_N)\,\mathrm{E}_{N,1}[P_2-P_1] + \Gamma_N \mathrm{E}_{N,1}[X-P_1] + cq_0}{a\Omega_N + c\left(1+\left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2+c}\right)^2\right)}
$$

where

$$
\begin{aligned}
\Omega_N = \quad & \left(\frac{c}{a\mathrm{Var}_{N,2}[X]+c} + \beta_3\left(\frac{a\mathrm{Var}_{N,2}[X]}{a\mathrm{Var}_{N,2}[X]+c}\right)^2\right)^2 \sigma_{e_1}^2 + \left(\frac{c}{a\mathrm{Var}_{N,2}[X]+c} + \beta_4\left(\frac{a\mathrm{Var}_{N,2}[X]}{a\mathrm{Var}_{N,2}[X]+c}\right)^2\right)^2 \sigma_{e_2}^2 \\
& + \left(1-\beta_3\left(\frac{a\mathrm{Var}_{N,2}[X]}{a\mathrm{Var}_{N,2}[X]+c}\right)^2\right)^2 \sigma_{\Delta e_1}^2 + \left(1-\beta_4\left(\frac{a\mathrm{Var}_{N,2}[X]}{a\mathrm{Var}_{N,2}[X]+c}\right)^2\right)^2 \sigma_{\Delta e_2}^2 \\
& + \left(1-\left(\frac{a\mathrm{Var}_{N,2}[X]}{a\mathrm{Var}_{N,2}[X]+c}\right)^2\right)^2 \beta_{\nu_2}^2\sigma_{\nu_2}^2 + \left\{\frac{c}{a\mathrm{Var}_{N,2}[X]+c}\right\}^2 \sigma_\varepsilon^2
\end{aligned}
$$

and

$$
\begin{aligned}
\Gamma_N \;=\;& \frac{c}{a\mathrm{Var}_{N,2}[X]+c}+c\frac{a\mathrm{Var}_{N,2}[X]}{(a\mathrm{Var}_{N,2}[X]+c)^2} \\
&+a\left\{\frac{\left(c\,(a\mathrm{Var}_{N,2}[X]+c)+\beta_3\,(a\mathrm{Var}_{N,2}[X])^2\right)(a\mathrm{Var}_{N,2}[X]+c-\beta_3\,(2a\mathrm{Var}_{N,2}[X]+c))}{(a\mathrm{Var}_{N,2}[X]+c)^4}\right\}\sigma_{e_1}^2 \\
&+a\left\{\frac{\left(c\,(a\mathrm{Var}_{N,2}[X]+c)+\beta_4\,(a\mathrm{Var}_{N,2}[X])^2\right)(a\mathrm{Var}_{N,2}[X]+c-\beta_4\,(2a\mathrm{Var}_{N,2}[X]+c))}{(a\mathrm{Var}_{N,2}[X]+c)^4}\right\}\sigma_{e2}^2 \\
&+a\left\{\left(1-\beta_3\left(\frac{a\mathrm{Var}_{N,2}[X]}{a\mathrm{Var}_{N,2}[X]+c}\right)^2\right)\left(\frac{(1-\beta_3)\,(2a\mathrm{Var}_{N,2}[X]+c)}{(a\mathrm{Var}_{N,2}[X]+c)^2}\right)\right\}\sigma_{\Delta e_1}^2 \\
&+a\left\{\left(1-\beta_4\left(\frac{a\mathrm{Var}_{N,2}[X]}{a\mathrm{Var}_{N,2}[X]+c}\right)^2\right)\left(\frac{(1-\beta_4)\,(2a\mathrm{Var}_{N,2}[X]+c)}{(a\mathrm{Var}_{N,2}[X]+c)^2}\right)\right\}\sigma_{\Delta e_2}^2 \\
&-a\left\{\left(1-\left(\frac{a\mathrm{Var}_{N,2}[X]}{a\mathrm{Var}_{N,2}[X]+c}\right)^2\right)\left(\frac{2a\mathrm{Var}_{N,2}[X]+c}{(a\mathrm{Var}_{N,2}[X]+c)^2}\right)\right\}\beta_{\nu_2}^2\sigma_{\nu_2}^2 \\
&+a\left\{\frac{c}{(a\mathrm{Var}_{N,2}[X]+c)^2}\right\}^2\sigma_\varepsilon^2
\end{aligned}
$$

B Confirming the linearity assumption

In equations (8) and (9), it was convenient to assume the market price to be a linear function of the state variables. Using the derived investor demand functions, we can verify this is true using the market clearing conditions.

In period 1, the market clears when

$$Q_0=\lambda_N q_{N,1}+\lambda_S q_{S,1}+\lambda_L q_{L,1}+\frac{\nu_1}{a\sigma_\varepsilon^2+c}.$$

The demand functions for the short-horizon and long-horizon investors are both linear in $\mathrm{E}[X]$ and hence linear in the state variables, so substituting them into the market clearing condition shows the price to be linear in the state variables. The expectations of the risky payout will all be linear in

P_1, which can be seen from substituting in the demand functions to the market clearing condition

$$\begin{aligned} P_1 \quad \propto \quad & \left\{ \lambda_S \left(\frac{(1-\Gamma_S)\,\beta_3 + \Gamma_S}{a\Omega_S + c\left(1 + \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2+c}\right)^2\right)} \right) + \lambda_L \left(\frac{(1-\Gamma)\,\beta_3\theta_1 + \Gamma\theta_1}{a\Omega + c\left(1 - \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2+c}\right)^2\right)} \right) \right\} \theta_1 \\ & \left\{ \lambda_L \left(\frac{(1-\Gamma)\,\beta_4 + \Gamma}{a\Omega + c\left(1 - \left(\frac{a\sigma_\varepsilon^2}{a\sigma_\varepsilon^2+c}\right)^2\right)} \right) \right\} \theta_2 \\ & + \left\{ \frac{1}{a\sigma_\varepsilon^2 + c} \right\} \nu_1 \end{aligned}$$

This confirms (8).

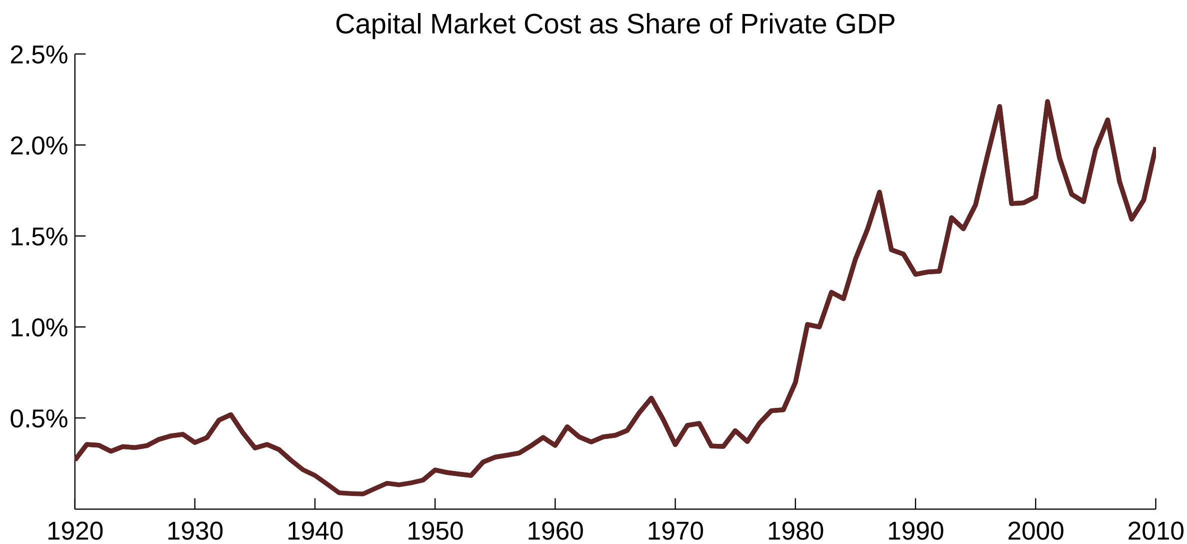
Similarly, in period 2 the market clearing condition shows that

$$P_2 \propto \frac{\lambda_S + \lambda_L}{a\sigma_\varepsilon^2 + c}\,(\theta_1 + \theta_2) + \left\{ \frac{1}{a\sigma_\varepsilon^2 + c} \right\} \nu_2,$$

which confirms (9).

Figure 1: Capital Market Spending and Compensation

The upper plot shows the share of GDP attributed to the capital markets sector using the gross value added measure, and the lower plot shows the ratio of average employee compensation in the capital markets sector relative to the US private industry average. The primary source for these calculations is the industry accounts data published by the US Bureau of Economic Analysis as of March 2011. Capital markets-related industries are described in Table 1. Data prior to 1947 comes from Philippon (2012).



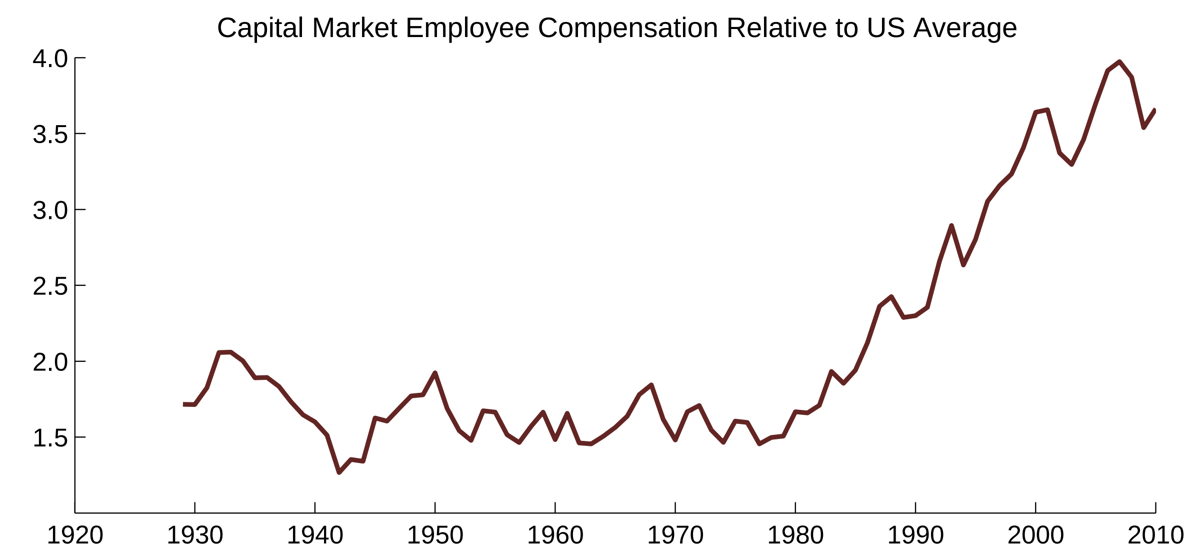
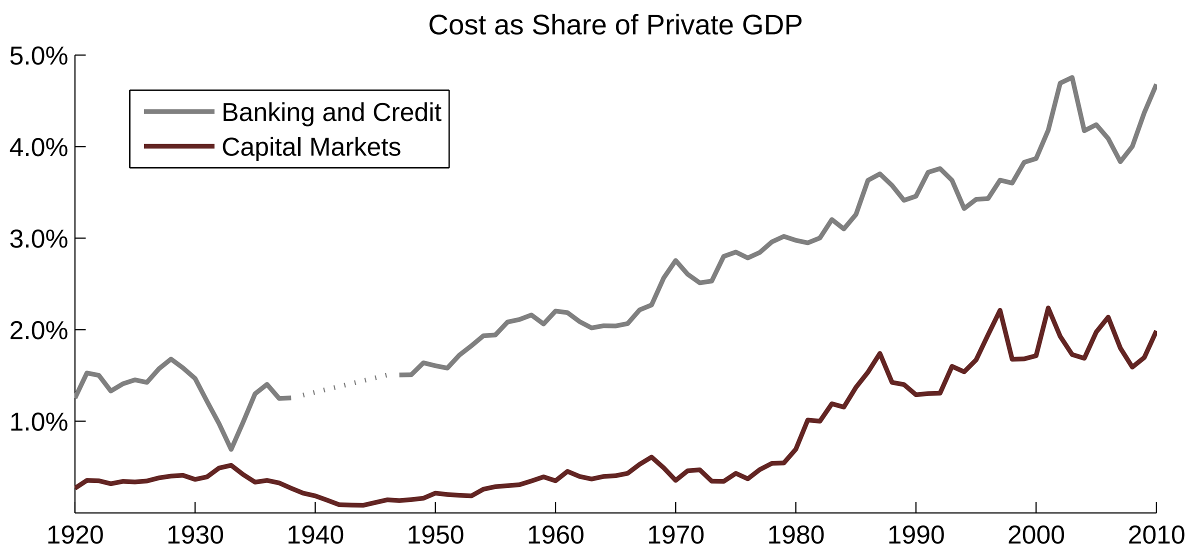


Figure 2: Contrasting Banking and Credit vs. Capital Market Activities

The upper plot contrasts the cost of banking and credit activity with the cost of capital markets using gross value added, and the lower plot shows the respective employee compensation ratios relative to the US private industry average. The primary source for these calculations is the industry accounts data published by the US Bureau of Economic Analysis as of March 2011. The classification to industry groups is shown in Table 1. Data prior to 1947 comes from Philippon (2012).



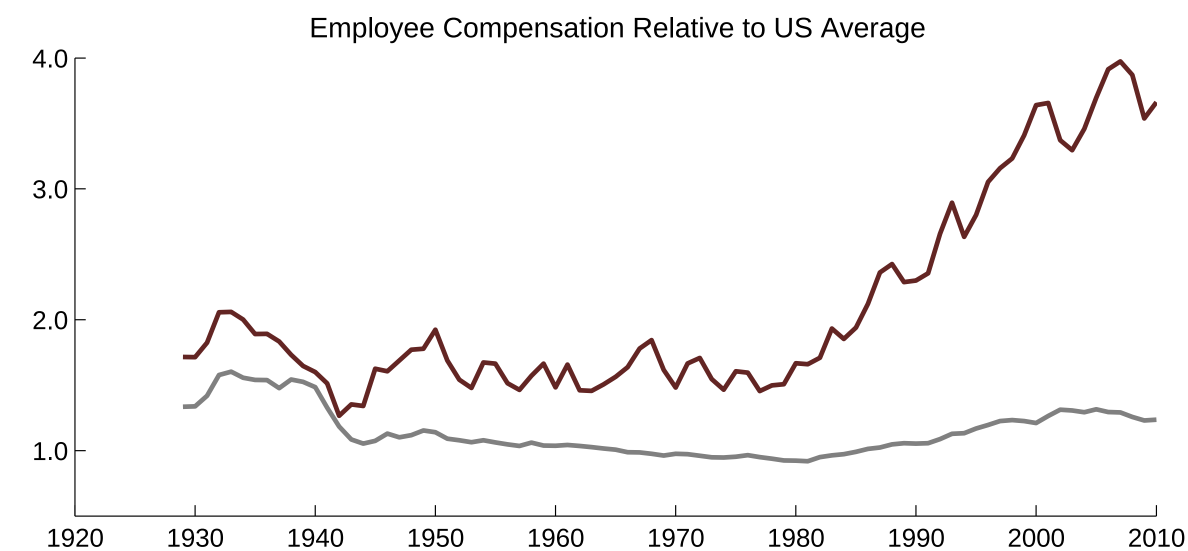


Figure 3: Intuition behind model equilibrium

The plots below correspond to the model presented in the paper in the one-period setting, $T = 1$. The model parameters are: $Q_0 = 1$, $\bar{X} = 100$, $\sigma_\epsilon^2 = \sigma_\theta^2 = 10^2$, $\sigma_\nu^2 = 2^2$, $a = 0.1$, $c = 10$, and $k = 1$. For illustration, the investment supply is allowed to be elastic in the short-run ($\Delta Q = b(P - \mathrm{E}[P])$), with linear supply parameter $b = 0.2$. The left axis plots the expected utility for the informed speculators and the uninformed passive investors. The right axis plots how the distribution of the market price, P, changes with respect to the quantity of informed speculators.

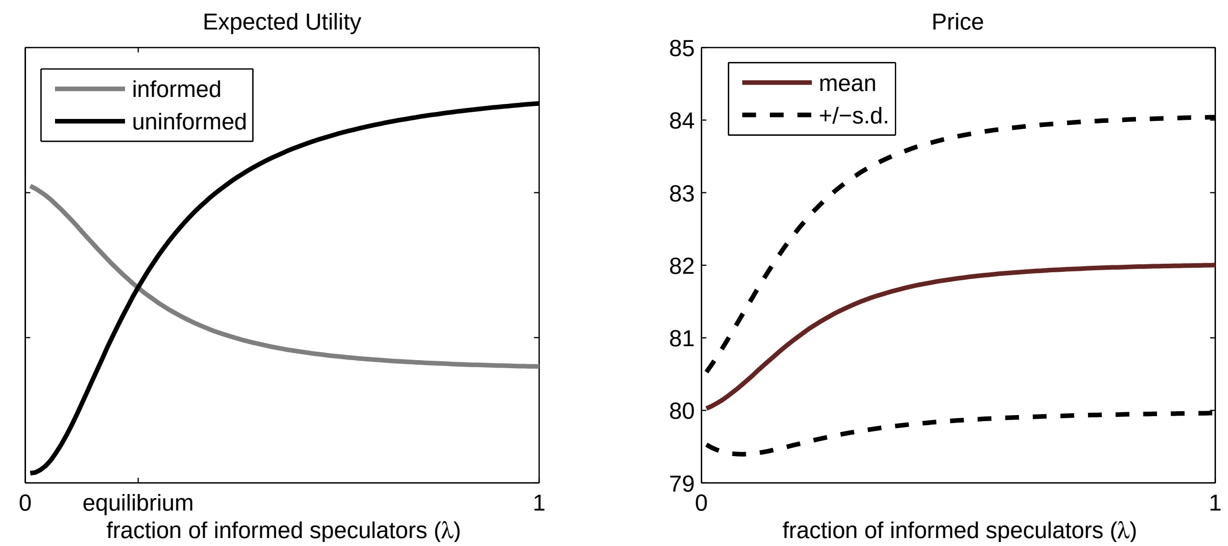


Figure 4: Relationship between transaction costs and active investing

The plots below graph the effect of transaction costs (c) on the equilibrium quantity of active investing and capital market spending. To keep the illustration simple, the one-period setting, $T = 1$ of the model is used with parameters: $Q_0 = 1$, $\bar{X} = 100$, $\sigma_\epsilon^2 = \sigma_\theta^2 = 10^2$, $\sigma_\nu^2 = 2^2$, $a = 0.1$, and $k = 1$. For illustration, the investment supply is allowed to be elastic in the short-run ($\Delta Q = b(P - \mathrm{E}[P])$), with linear supply parameter $b = 0.2$. The left axis plots the equilibrium quantity of informed speculators (λ) as a function of the exogenous transaction costs parameter, c. The right axis plots the quantity of resources spent on active management ($\lambda \times k$) as well as total capital market spending, which includes trading costs.

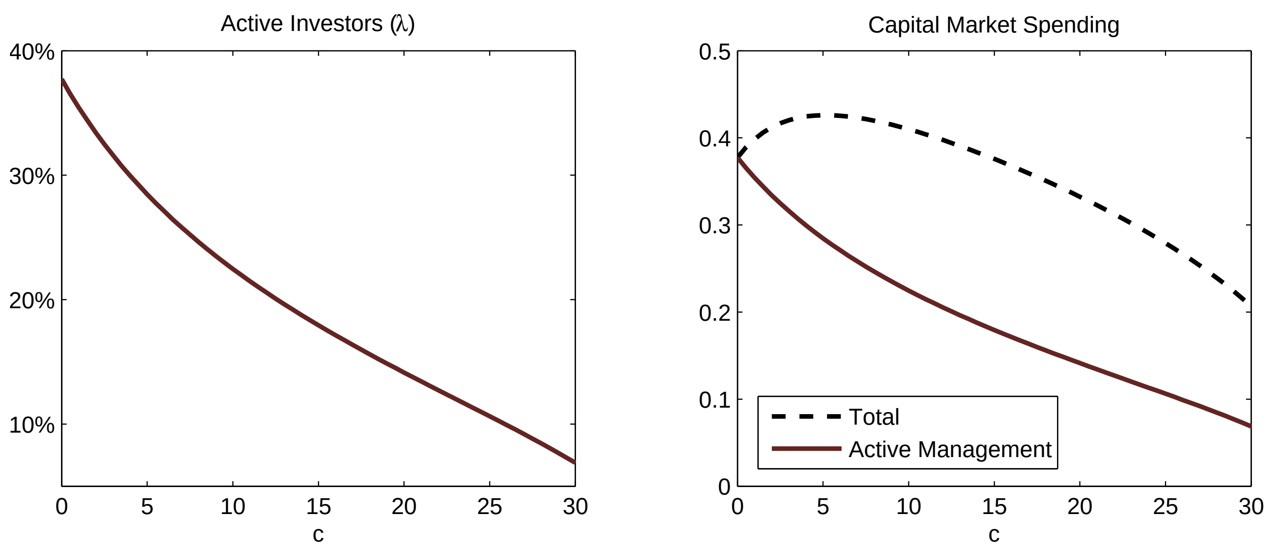


Figure 5: NYSE Commission Schedule, 1956

The image below shows the New York Stock Exchange minimum commission schedule for 1956, as reported on page 7 of the NYSE Fact Book for 1965.

MEMBERS

MINIMUM COMMISSION RATES

Round Lots

A simplified method for determining minimum commission charges by member firms of the New York Stock Exchange for 100 share round-lot purchases and sales of New York Stock Exchange listed stocks is shown below. (For orders involving multiples of 100 shares, e.g., 200, 300, etc., multiply the 100 share commission by 2, 3, etc.)

Price per Share	Minimum Commission Rate per 100 Shares
Between **\$1** and **\$20**	**\$5** plus price per share
From **\$20** to **\$50**	**\$15** plus 1/2 price per share
Between **\$50** and **\$150**	**\$35** plus 1/10 price per share
At **\$150** or more	**\$50**

EXAMPLES

of Minimum Commission Charges for 100 Shares

At \$15 a share, the commission would be \$5 plus \$15 or **\$20.**

At \$35 a share, the commission would be \$15 plus \$17.50 or **\$32.50.**

At \$100 per share, the commission would be \$35 plus \$10 or **\$45.**

At \$200 per share, the commission would be **\$50.**

Figure 6: Transaction Cost Time Series

The figure below plots the composite transaction cost measure, constructed as described in section 3.1, plotted alongside three other comparison series. The orange series shows the cost of trading a stock with a nominal share price of $30 according to the published NYSE commission schedule, the red series shows the decrease in commission costs as measured by the Securities and Exchange Commission in their analyses of the effects of commission deregulation, and the green line plots the average equity commission charge collected in a survey of institutional investors by Greenwich Associates, a financial consulting firm.

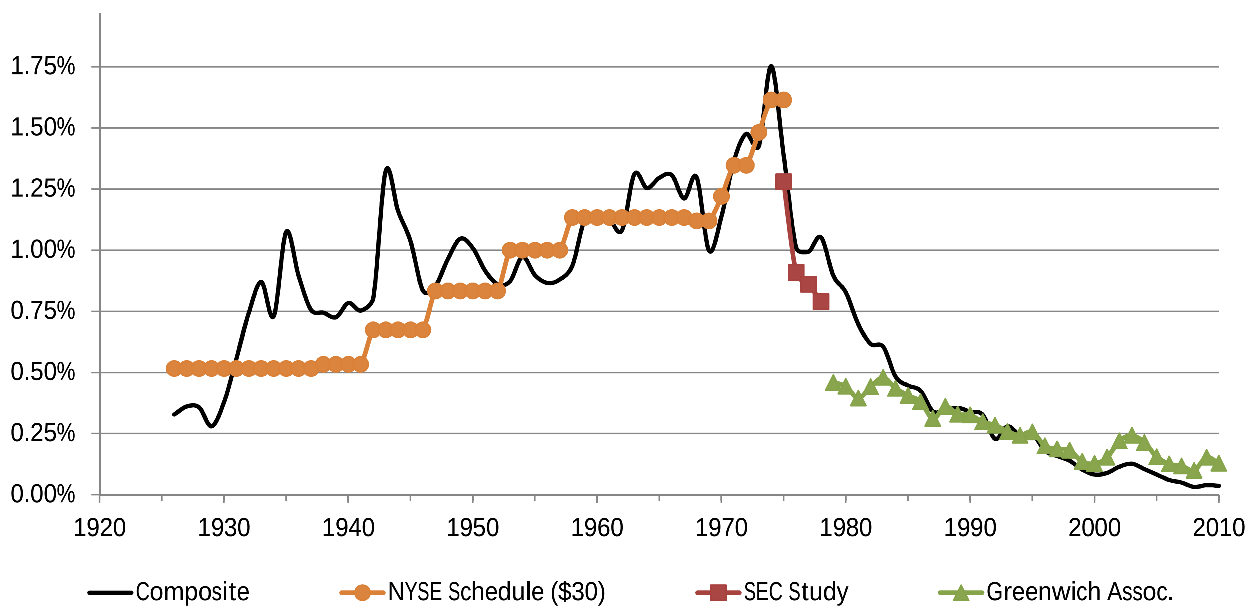


Figure 7: Round Trip Trading Costs, May 1974

The figure below plots the commission schedule in effect in May of 1974, one year before deregulation, with the red line showing the explicit commission charge and the black line illustrating the additional effect of paying half of the bid-ask spread, commonly 1/4 of a dollar. These are plotted on top of a histogram representing the distribution of nominal share prices at the time, as reported by CRSP.

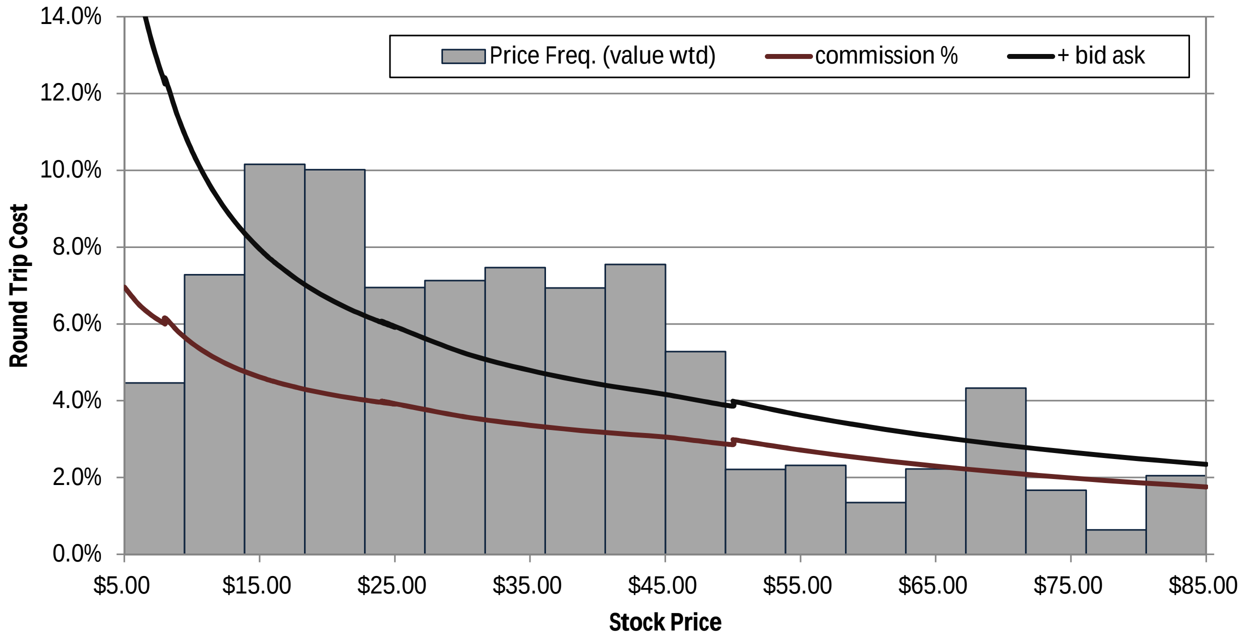


Figure 8: Predicting the cost of capital markets using the cost of transacting

The figure below plots in red the percentage of national income consumed by capital markets related activity using a GDP value-added measure divided by private GDP calculated using data from the Bureau of Economic Analysis. The dotted line shows the fit of a time series regression using the composite commission time series and a linear time trend.

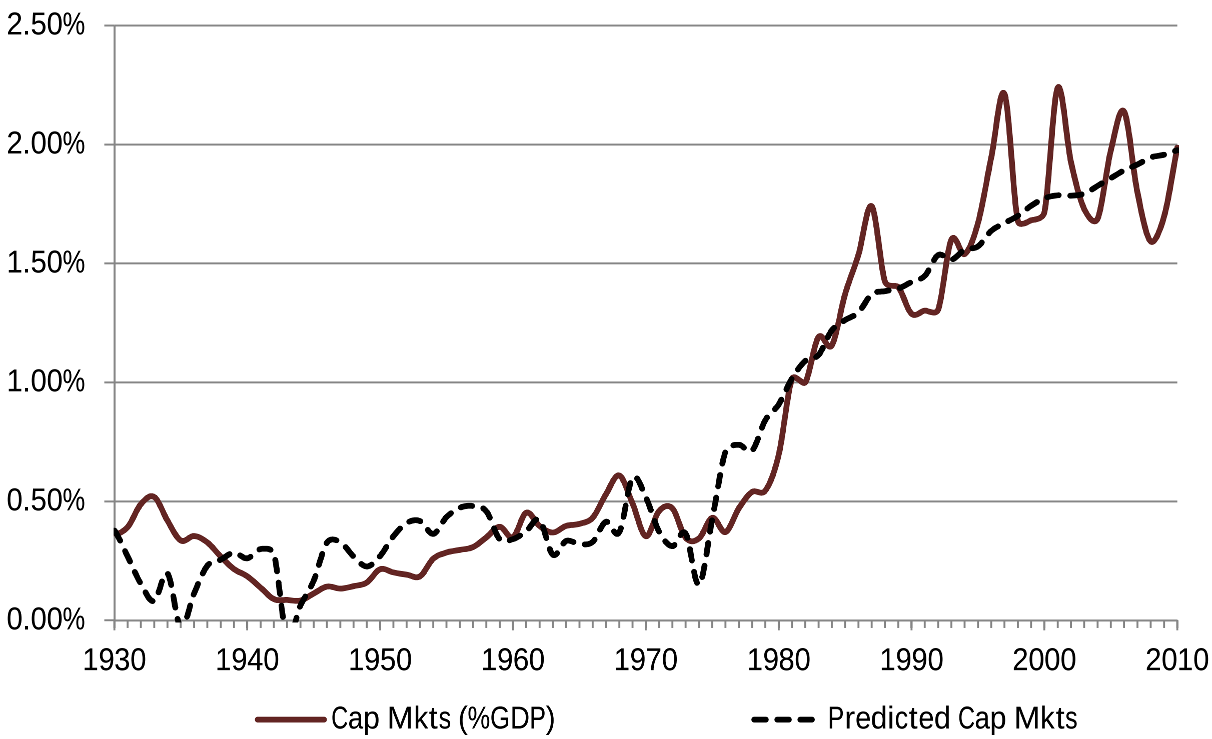


Figure 9: Rolling Regression Coefficient and Moving Average, 1965-2010

The two axes plot the results of the rolling regressions described in section 4.3. The top axis plots the estimated regression coefficients and the lower axis plots the square root of the mean squared error (RMSE) and the R^2 values.

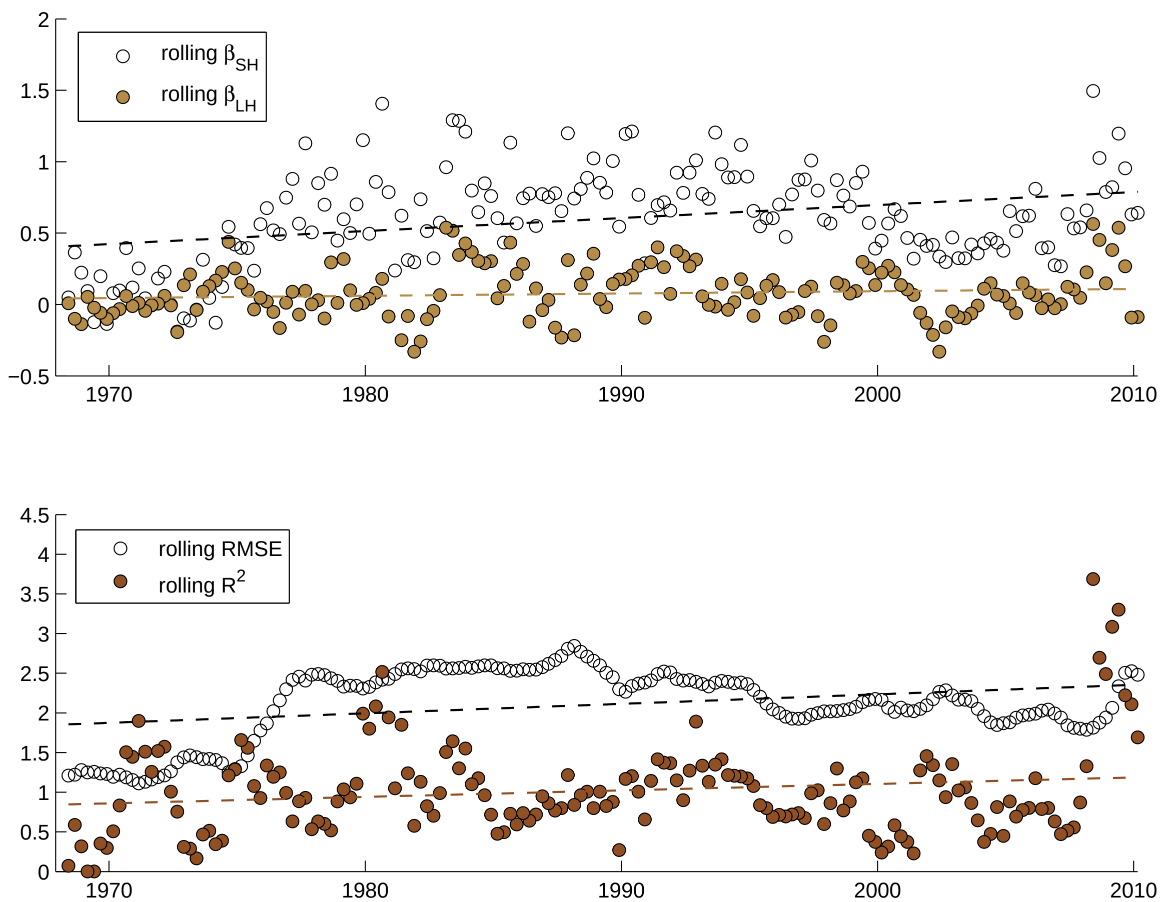


Figure 10: Welfare illustration in the case of inelastic investment supply (short-horizon)

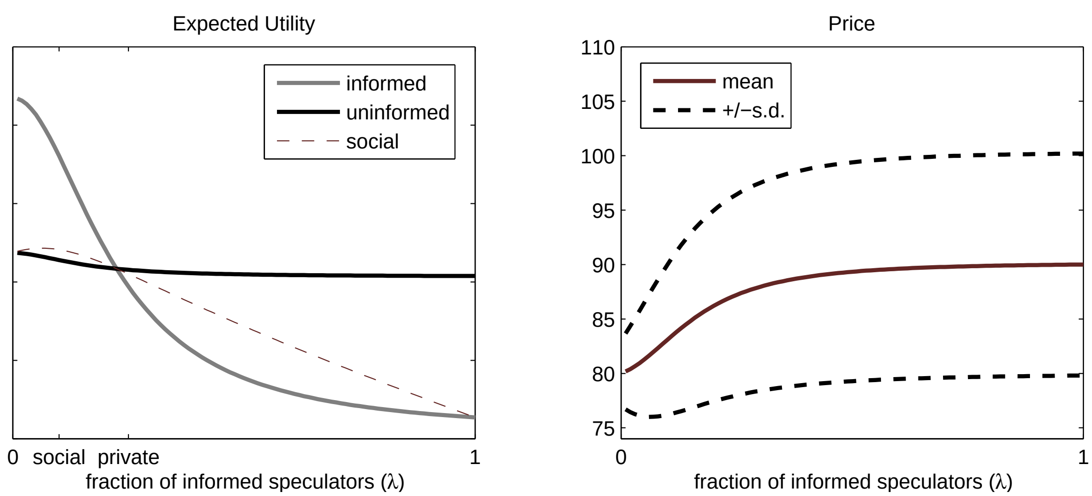


Figure 11: Welfare illustration in the case of elastic investment supply (long-horizon)

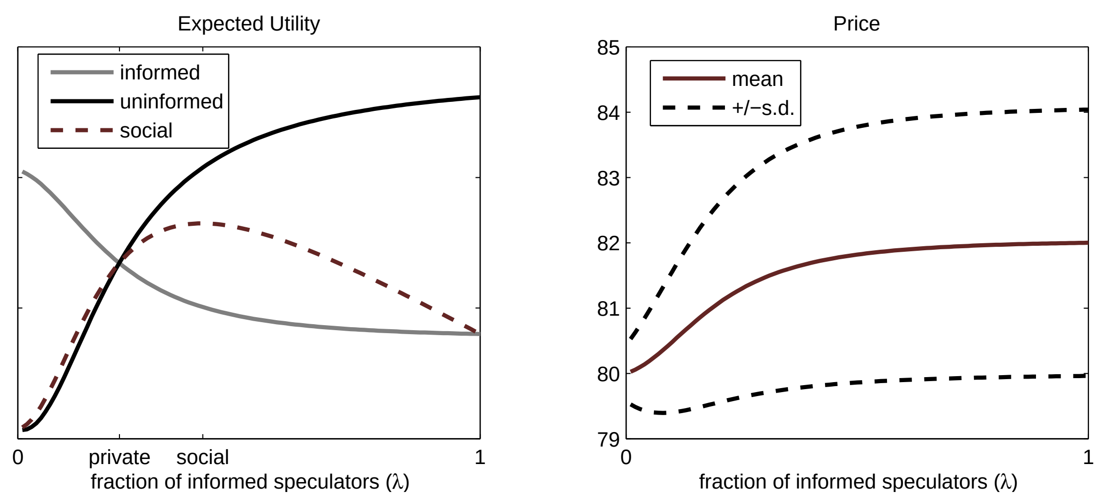


Table 1: Financial sector components in national income accounts

This table shows the components of the financial sector and the associated NAICS codes as used by the US Bureau of Economic Analysis in their national income accounts. The grouping of the components has not always been historically consistent. The highlighted industries are those which will be termed the capital markets sector and are the primary focus of this paper.

Finance, Insurance, and Real Estate
Banking and Credit (521 & 522)
Banking
Credit agencies other than banks
Capital Markets (523 & 525)
Security and commodity brokers
Funds, trusts, and other financial vehicles
Holding and other investment offices
Insurance (524)
Insurance carriers
Insurance agents, brokers, and service
Real Estate and Leasing (531, 532, 533)
Real Estate
Rental and leasing services and lessors of intangible assets

Table 2: Time series summary statistics and correlations

This table shows summary statistics for annual data on: the average commission cost of transacting stocks in the United States (tcost) constructed as described in section 3.1; the percentage of national income consumed by capital markets related activity using a GDP value-added measure divided by private GDP calculated using data from the Bureau of Economic Analysis (capmkt%); the ratio of the average salary for employees in capital markets related industries relative to the average salary across all private-sector employees using data from the Bureau of Economic Analysis (comp ratio); and the annual turnover in US equities measured by dividing annual volume by shares outstanding as reported in CRSP. Annual observations are used over the period 1927-2010 to calculate the mean, standard deviation and various percentiles in the upper panel. Correlations are displayed in the lower panel.

1927-2010	mean	std.	1 %ile	50 %ile	99 %ile
tcost (bps)	71.1	43.6	3.6	78.4	152.0
capmkt% (bps)	78.8	65.7	8.5	43.1	221.6
comp ratio	2.09	0.77	1.20	1.72	3.92
turnover	55.7	58.9	7.3	30.4	277.1
Correlation	tcost	capmkt	comp	turnover	
tcost (bps)	1.00	-0.81	-0.83	-0.76	
capmkt (bps)	-0.81	1.00	0.90	0.72	
comp ratio	-0.83	0.90	1.00	0.87	
turnover	-0.76	0.72	0.87	1.00	

Table 3: Time series regressions of first differences

This table shows the results of regressing changes in the income share of capital markets (Δcapmkt), capital market compensation (Δcomp), and equity turnover by volume (Δturnover) on changes in the commission cost of stock transactions (Δtcost) with up to four lags. Newey-West adjusted t-statistics, with four lags, are reported in parentheses. Statistical significance is noted with: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	Δcapmkt (1)	Δcomp (2)	Δturnover (3)
Δtcost	-3.46	4.33	0.62
	(4.93)	(8.45)	(6.34)
$L(\Delta\text{tcost})$	-3.01	3.42	-4.24
	(6.20)	(9.77)	(5.55)
$L^2(\Delta\text{tcost})$	-12.98*	2.62	-2.95
	(7.29)	(11.67)	(5.64)
$L^3(\Delta\text{tcost})$	-2.41	-18.11**	-9.16**
	(4.77)	(7.07)	(4.13)
$L^4(\Delta\text{tcost})$	-6.06	-7.20	-7.34
	(6.56)	(7.95)	(5.86)
Constant	1.92	2.38	2.16
	(1.17)	(1.56)	(1.72)
Observations	80	80	80

Table 4: Summary Statistics for Panel Data Analysis

The summary statistics below are for the quarterly panel data collected for the 1,000 firms in the annual universe being analyzed. The universe is reset each year, taking the 1,000 largest firms by market cap. The first panel cover the full sample period, while the lower panel covers the 5-year window before fixed exchange regime was ended on May 1, 1975 up until 5-years after May 1, 1977—the date at which none of the collected series overlap with the fixed-rate commission regime.

	mean	std.	1 %ile	50 %ile	99 %ile
1966 - 2010					
Δx_t	-0.01	2.08	-9.88	0.05	8.83
price	104.10	23.56	6.24	32.50	132.60
turnover	2.36	3.20	0.10	1.39	14.58
r_L	0.124	0.446	-0.835	0.082	1.526
r_S	0.013	0.216	-0.577	0.011	0.598
r_A	0.006	0.156	-0.411	0.004	0.434
					(N = 134,128)
1970-1982					
Δx_t	0.02	1.97	-7.60	0.07	6.90
price	32.17	22.26	6.75	27.38	111.80
turnover	0.81	0.87	0.05	0.58	4.19
r_L	0.102	0.388	-0.785	0.075	1.245
r_S	0.017	0.193	-0.485	0.012	0.546
r_A	0.007	0.139	-0.336	0.004	0.394
					(N = 36,174)

Table 5: Base panel regression with time trend

The regression estimates below are the result of panel regressions of earnings news (Δx defined in section 4 of the paper) on past log returns, log returns interacted with a time trend. The regression also includes a constant term and constant *trend* variable, but the coefficients are not reported. Industry-clustered, heteroskedasticity robust standard errors are in parentheses below each estimated coefficient. Statistical significance is noted with: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1)	(2)	(3)
r_L	0.033	-0.001	0.029
	(0.028)	(0.026)	(0.053)
$r_L \times trend$			0.0000
			(0.0020)
r_S	0.667***	0.712***	0.315***
	(0.078)	(0.073)	(0.100)
$r_S \times trend$			0.0110***
			(0.0026)
r_A	0.720***	0.816***	1.380***
	(0.080)	(0.073)	(0.145)
$r_A \times trend$			-0.0208***
			(0.0056)
Fixed Effects			
# firms	3,061		3,061
# industries		66	
# quarters		175	
Observations	134,128	134,128	134,128

Table 6: Testing the May Day effect in the time series

The regression estimates below are the result of panel regressions of earnings news (Δx defined in section 4 of the paper) on past log returns and log returns interacted with a post-1975 dummy variable. Coefficients for constant term and constant post-1975 dummy are estimated but not reported. Industry-clustered, heteroskedasticity-robust standard errors are reported in parentheses. Statistical significance is noted with: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	full-sample (1)	10 yr window (2)	5 yr window (3)	3 yr window (4)
r_L	0.010 (0.035)	0.010 (0.035)	0.016 (0.051)	0.069 (0.074)
$r_L \times post75$	0.031 (0.043)	0.078 (0.055)	-0.000 (0.062)	-0.137 (0.105)
r_S	0.234*** (0.066)	0.235*** (0.066)	0.255*** (0.086)	0.375** (0.153)
$r_S \times post75$	0.513*** (0.119)	0.560*** (0.180)	0.469** (0.208)	0.407 (0.319)
r_A	0.811*** (0.128)	0.812*** (0.129)	0.933*** (0.178)	0.870*** (0.251)
$r_A \times post75$	-0.124 (0.173)	0.514*** (0.191)	0.704** (0.287)	1.077** (0.496)
Fixed Effects				
# firms	3,058	1,653	1,205	1,059
Observations	128,114	55,184	30,160	18,070

Table 7: Testing May Day effect in the cross-section

Coefficients for constant term and unique permutations of constant dummies are not reported. Industry-clustered, heteroskedasticity-robust standard errors are reported in parentheses. Statistical significance is noted with: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	10 yr window (1)	5 yr window (2)	3 yr window (3)
Long-horizon return			
$R_{LH} \ldots$	-0.014	0.004	0.015
	(0.020)	(0.029)	(0.042)
$\times lowP$	0.025	-0.009	-0.023
	(0.120)	(0.137)	(0.180)
$\times midP$	-0.040	-0.087	-0.164
	(0.038)	(0.056)	(0.111)
$\times lowP \times post75$	0.137	-0.091	-0.193
	(0.191)	(0.201)	(0.261)
$\times midP \times post75$	0.125***	0.0639	0.113
	(0.047)	(0.062)	(0.135)
$\times highP \times post75$	-0.013	-0.040	-0.0659
	(0.191)	(0.287)	(0.496)
Short-horizon return			
$R_{SH} \ldots$	0.186***	0.225***	0.313***
	(0.040)	(0.058)	(0.089)
$\times lowP$	-0.0217	-0.027	-0.094
	(0.225)	(0.244)	(0.325)
$\times midP$	0.131	-0.0375	-0.025
	(0.146)	(0.168)	(0.209)
$\times lowP \times post75$	2.201***	2.056***	1.939***
	(0.416)	(0.520)	(0.675)
$\times midP \times post75$	0.292	0.368	0.067
	(0.241)	(0.248)	(0.292)
$\times highP \times post75$	0.164*	0.144*	0.170
	(0.082)	(0.084)	(0.121)
Fixed effects			
# industries	64	61	61
Observations	61,198	36,174	24,084

References

Bai, Jennie, Thomas Philippon and Alexi Savov. 2012. "Have Financial Markets Become More Informative?," *Working Paper, New York University.*

Berkowitz, Stephen A., Dennis E. Logue and Eugene A. Noser. 1988. "The Total Cost of Transactions on the NYSE.," *Journal of Finance*, (1), 97–112.

Bolton, Patrick, Tano Santos and Jose A Scheinkman. 2011. "Cream skimming in financial markets," *Working Paper, Columbia University.*

Brainard, William C. and James Tobin. "Asset markets and the cost of capital," in "Economic Progress, Private Values and Public Policy: Essays in Honor of william Fellner," Amsterdam: North, 1977. pp. 235–62.

Campbell, John Y and Kenneth A Froot. "International Experiences with Securities Transaction Taxes," in Jeffrey Frankel, ed., *The Internationalization of Equity Markets*, number January, Chicago, IL: University of Chicago Press, 1994. chapter 6, pp. 277–308.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers and Robert J. Waldmann. 1989. "The Size and Incidence of the Losses from Noise Trading," *The Journal of Finance*, *44* (3), 16.

Francis, Jennifer, Katherine Schipper and Linda Vincent. August 2002. "Earnings Announcements and Competing Information," *Journal of Accounting and Economics*, *33* (3), 313–342.

French, Kenneth R. 2008. "Presidential Address: The Cost of Active Investing," *The Journal of Finance*, *63* (4), 1537–1573.

Glode, Vincent, Richard C. Green and Richard Lowery. 2012. "Financial Expertise as an Arms Race," *Journal of Finance*, (forthcoming).

Greenwood, Robin and David Scharfstein. 2012. "The Growth of Modern Finance," *Working Paper, Harvard Business School.*

Grossman, Sanford J. and Joseph E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets," *The American Economic Review*, *70* (3), 393 – 408.

Hirshleifer, Jack. 1971. “The Private and Social Value of Information and the Reward to Inventive Activity.,” *American Economic Review*, *61* (4), 561–574.

Murphy, Kevin M., Andrei Shleifer and Robert W Vishny. 1991. “The allocation of talent:,” *Quarterly Journal of Economics*, (2), 503–530.

Pástor, Luboš and Robert F Stambaugh. 2010. “On the Size of the Active Management Industry.,” *Working Papers – Financial Institutions Center at The Wharton School.*

Philippon, Thomas. 2010. “Financiers versus Engineers: Should the Financial Sector be Taxed or Subsidized?,” *American Economic Journal: Macroeconomics*, *2* (3), 25.

_ . 2012. “Has the US Finance Industry Become Less Efficient?,” *Working Paper, New York University*, (May).

_ and Ariell Reshef. 2013. “Wages and Human Capital in the U . S . Financial Industry : 1909-2006,” *Quarterly Journal of Economics.*

Simsek, Alp. 2012. “Speculation and Risk Sharing with New Financial Assets,” *Working Paper, Harvard University.*

Stein, Jeremy C. 1987. “Informational Externalities and Welfare-Reducing Speculation,” *Journal of Political Economy*, *95* (6), 1123–1145.

Summers, Lawrence H. and Victoria P. Summers. 1989. “When financial markets work too well: A cautious case for a securities transactions tax,” *Journal of Financial Services Research*, *3* (2-3), 261–286.

Tinic, Seha M. and Richard R. West. 1980. “The securities industry under negotiated brokerage commissions:,” *The Bell Journal of Economics*, *11* (1), 29–41.

Weld, William C, Roni Michaely, Richard H Thaler and Shlomo Benartzi. 2009. “The Nominal Share Price Puzzle,” *Journal of Economic Perspectives*, *23* (2), 121–142.