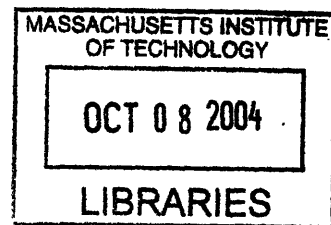


Three Essays in Financial Economics

by

Mark M. Westerfield



Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of **ARCHIVES**
Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2004

© Mark M. Westerfield, MMIV. All rights reserved.

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

Author
Department of Economics
September 1, 2004

Certified by
Daron Acemoglu
Charles P. Kindleberger Professor of Applied Economics
Thesis Supervisor

Accepted by
Peter Temin
Elisha Gray II Professor of Economics
Chairman, Departmental Committee on Graduate Studies

Three Essays in Financial Economics

by

Mark M. Westerfield

Submitted to the Department of Economics
on September 1, 2004, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

The Price Impact and Survival of Irrational Traders

with Leonid Kogan, Stephen Ross, and Jiang Wang

Milton Friedman argued that irrational traders will consistently lose money, won't survive and, therefore, cannot influence long run equilibrium asset prices. Since his work, survival and price impact have been assumed to be the same. In this paper, we demonstrate that survival and price impact are two independent concepts. The price impact of irrational traders does not rely on their long-run survival and they can have a significant impact on asset prices even when their wealth becomes negligible. We also show that irrational traders' portfolio policies can deviate from their limits long after the price process approaches its long-run limit. We show, in contrast to a partial equilibrium analysis, these general equilibrium considerations matter for the irrational traders' long-run survival. In sum, we explicitly show that price impact can persist whether or not the irrational traders survive.

Market Composition and Equity Market Formation

I present a model of agents with heterogeneous beliefs who must choose whether to participate in an asset market. Investor composition affects asset prices, so the participation choice creates an externality: agents premise their entry decisions and asset valuations on the participation decisions of other agents. Investment banks can use their pricing discretion to change agents' participation decisions in IPOs. When combined with the effect of investor composition on price, this implies that allocation procedures with pricing discretion will dominate open auctions as a means to market securities. In a model with noise and rational traders, I show that noise trader participation will lower the expected value of a stock in the aftermarket, so firms will desire to exclude them from their IPOs. The "money left on the table" due to underpricing in the IPO allocation is not capital the firm could have raised; instead, it is the empirical regularity associated with obtaining a high quality aftermarket, high equity valuation, and higher proceeds to the issuer.

Heterogeneous Beliefs and the Principal-Agent Problem

with Tobias Adrian

We examine the principal-agent problem in a simple continuous time framework when potential agents have heterogeneous priors. We find that the principal prefers agents with priors very different from his own. The principal will create a contract that includes side-bets to exploit gains from trade created by heterogeneous priors despite the distortionary effect on effort choice. In a semi-dynamic labor market, the principal can optimally choose to churn his employees to prevent them from learning about project profitability, even when agents' skills are increasing with job tenure. We develop several empirical predictions, and relate our model to the labor market in the financial industry.

Thesis Supervisor: Daron Acemoglu

Title: Charles P. Kindleberger Professor of Applied Economics

Acknowledgments

To my advisors, Daron Acemoglu, Leonid Kogan, Stephen Ross, and Jiang Wang, I thank you for all the time you put in and for all the support you gave me. Leonid, I thank you especially, because for all the pressure on you as junior faculty, you were still willing to go through ideas and formulations with me in great detail. Daron, I thank you for helping me keep my perspective both broad and deep and for spending the time you did. To Steve and Jiang, you performed the role of senior advisors to perfection – knowledge and history you have in spades. I also thank the other MIT faculty that helped me on my way; they are far too numerous to mention. To my parents, maybe the only graduate student parents who understood what their child was doing: you were cheerleaders and advisors of the first order. Finally, to Tracey, my constant companion and now wife: we were together from the first day of orientation, and you were always an accepting and loving partner. My life would be so much the poorer without you and your light.

Contents

1	The Price Impact and Survival of Irrational Traders	9
1.1	Introduction	9
1.2	An Example	12
1.3	The Model	16
1.4	The Equilibrium	18
1.5	Logarithmic Preferences	20
1.6	Survival	27
1.7	The Price Impact of Irrational Traders	31
1.8	Portfolio Policies	33
1.9	Heuristic Partial Equilibrium Analysis of Survival	37
1.10	Conclusion	41
1.11	Appendix	42
2	Market Composition and Equity Market Formation	59
2.1	Introduction	59
2.2	A Model of the Aftermarket	65
2.2.1	Assets and Agents	65
2.2.2	The Stock Price and Expected Utility	71
2.2.3	Stock Market Participation	73
2.2.4	The Three Period Aftermarket	82
2.2.5	Discussion	86
2.3	The IPO Allocation	87

2.3.1	The Allocation Procedure and the Pre-Market	88
2.3.2	The Investor's Problem	92
2.3.3	The Firm's Problem	98
2.3.4	Discussion	102
2.4	Extension to an OLG Aftermarket	103
2.4.1	The OLG Aftermarket	104
2.4.2	The Firm's Problem	110
2.4.3	Discussion	114
2.5	Conclusion	115
2.6	Appendix	116
3	Heterogeneous Beliefs and the Principal-Agent Problem	127
3.1	Introduction	127
3.2	The Model	132
3.2.1	Assumptions	132
3.2.2	Equilibrium	135
3.2.3	Discussion	143
3.3	The Labor Market	144
3.3.1	Assumptions	145
3.3.2	Equilibrium	146
3.4	Predictions and Discussion	149
3.4.1	Comparative Statics and Predictions	149
3.4.2	Comparison to a Rational Model	152
3.5	Conclusion	153
3.6	Appendix	154
3.6.1	A Discrete Time Analog	154
3.6.2	Proofs	156
3.6.3	The Kalman-Bucy Filter	158

Chapter 1

The Price Impact and Survival of Irrational Traders

Written with Leonid Kogan, Stephen Ross, and Jiang Wang.

1.1 Introduction

Most neoclassical asset pricing models rely on the assumption that market participants (traders) are rational in the sense that they behave in ways that are consistent with the objective probabilities of the states of the economy (e.g., Radner (1971) and Lucas (1978)). More particularly, they maximize expected utilities using the true probabilities of uncertain economic states. This approach is firmly rooted in the tradition of going from the normative to the positive in economics, yet there is mounting evidence that it is not descriptive of the observed behavior of the average market participant (see, e.g., Alpert and Raiffa (1959), Benartzi and Thaler (2001), Black (1986), Kahneman and Tversky (1979), and Odean 1998)). How the presence of traders with incorrect beliefs may affect the behavior of financial markets remains an open question.

It has long been argued (see, e.g., Friedman (1953)) that irrational traders who use wrong beliefs cannot survive in a competitive market. Trading under the wrong

beliefs causes them to lose their wealth. In the long-run, it is the rational traders who control most of the wealth and determine asset prices. Using a partial equilibrium model, De Long, Shleifer, Summers and Waldmann (1991) suggest that traders with wrong beliefs may survive in the long-run because they may hold portfolios with higher growth rates and therefore can eventually outgrow the rational traders.¹ In contrast, in a general equilibrium setting, Sandroni (2000) and Blume and Easley (2001) show that with intermediate consumption, irrational traders do not survive in the long-run.

The efficiency of financial markets is the principal motivation behind the interest in the survival of irrational traders. If irrational traders impact asset prices, then markets will not be efficient, either informationally or allocationally. Implicitly, the discussion on survival is based on the assumption that survival is a necessary condition for long-run price impact. It is thought that irrational traders have to control a significant amount of wealth in order to affect – or ‘infect’ – prices with their irrational beliefs. In this paper, we show that this assumption is false and that irrational traders can maintain a large price impact even as their relative wealth diminishes towards zero over time.

Our analysis is conducted with a parsimonious general equilibrium model inhabited by both rational traders and irrational traders. Traders only care about their terminal consumption. We are able to derive an explicit solution to the model and obtain conditions under which the irrational traders can survive in the long run in the sense that their share of the total wealth does not go to zero over time. However, we show that even when irrational traders do not survive, with a negligible amount of wealth they can still exert significant influence on the asset price over a long period of time.

Underlying this initially counterintuitive result is a solid economic intuition. Under incorrect beliefs, irrational traders express their views by taking positions (bets) on extremely unlikely states of the economy. As a result, the state prices of these ex-

¹See also Figlewski (1978) for a discussion on the notion of long-run survival.

treme states can be significantly affected by the beliefs of the irrational traders, even with negligible wealth. In turn, these states, even though highly unlikely, can have a large contribution to current asset prices. This is especially true for states associated with extremely low levels of aggregate consumption in which the traders' marginal utilities are very high and so too are state prices. The beliefs of the irrational traders on these low probability but high marginal utility states can influence current asset prices and their dynamics. Furthermore, irrational traders need not take extreme positions in order to influence prices. Our formal analysis clearly verifies this conceptual distinction between the long-run price impact and the long-run survival of irrational traders.

The possibility that irrational traders may have a significant price impact with a negligible share of wealth also has important implications for their survival. In the partial equilibrium analysis of De Long, Shleifer, Summers and Waldmann (1991) (DSSW, thereafter), it was assumed that when the irrational traders control only a negligible fraction of the total wealth, they have no impact on asset prices, i.e., asset prices behave as if the irrational traders are absent. Given the rationally determined prices, DSSW then show that the wealth of irrational traders can grow at a faster rate than the wealth of the rational traders, allowing the irrational traders to recover from their losses and survive in the long-run. Although such an argument is illuminating, it is based on unreliable premises. As we have argued, irrational traders may still influence prices with diminishing wealth. Moreover, such a possibility can significantly affect the irrational traders' portfolio policies in ways that make their recovery from losses difficult.

The paper proceeds as follows. In Section 1.2, we provide a simple example to illustrate the possibility for an agent to affect asset prices with a negligible wealth. Section 1.3 describes a canonical economy similar to that of Black and Scholes (1973), but in the presence of irrational traders who have persistently wrong beliefs about the economy, and Section 1.4 describes the general equilibrium of this economy. Section 1.5 treats the special case of logarithmic preferences and demonstrates that even

though irrational traders never survive in this case, they nevertheless can still influence long-run asset prices. Sections 1.6, 1.7, and 1.8 analyze the survival of irrational traders, their price impact and their portfolio policies for the case of risk aversion different from one. Section 1.9 discusses the importance of equilibrium effects on the survival of irrational traders. Section 1.10 concludes the paper with a short summary and some suggestions for future research. All proofs are given in the appendix.

1.2 An Example

We begin our analysis by considering a simple, static Arrow-Debreu economy and will show that an agent with only a negligible amount of wealth can have a significant impact on asset prices by using certain trading policies.

The economy has two dates, 0 and 1. It is endowed with one unit of a risky asset, which pays a dividend D only at date 1. The realization of D falls in $[0, 1]$ with probability density $p(D) = 2D$, which is plotted in Figure 1-1(a).

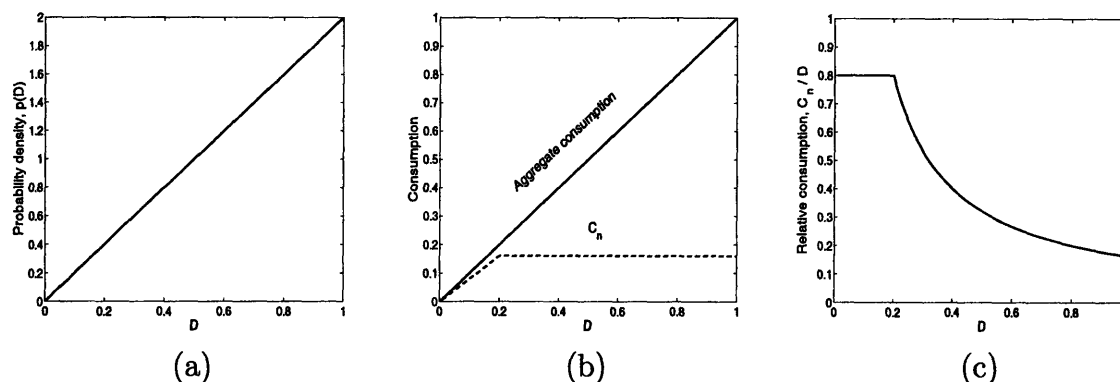


Figure 1-1: Probability distribution of the stock dividend (the left panel), the aggregate consumption level (D) and the noise trade consumption (C_n) when he is present (the middle panel), and the relative consumption of the noise trader (C_n/D , right panel). Parameter δ is set to 0.2.

There is a complete set of Arrow-Debreu securities traded in a competitive financial market at date 0. Shares of the stock and a risk-free bond with a sure payoff of 1 at date 1, both of which are baskets of the Arrow-Debreu securities, are also traded. We use the bond as the numeraire for the security prices at date 0. Thus, the bond

price is always 1.

We first consider the economy when it is populated by a representative agent with a logarithmic utility function over consumption at date 1, $u(C) = \log C$. It immediately follows that $C = D$ and the state price density, denoted by ϕ^* , is

$$\phi^*(D) = a^* u'(D) = \frac{a^*}{D}$$

where a^* is a constant. The price of any payoff X is then given by

$$P = E[X \cdot \phi^*].$$

In particular, the price of the bond is

$$B = E[1 \cdot \phi^*] = \int_0^1 \frac{a^*}{D} p(D) dD = \int_0^1 2a^* dD = 2a^* = 1$$

which gives $a^* = \frac{1}{2}$. The price of the stock is then given by

$$S^* = E[D \cdot \phi^*] = \int_0^1 D \cdot \frac{1}{2D} p(D) dD = \int_0^1 D dD = \frac{1}{2}.$$

Now we introduce another trader to the economy who has a negligible amount of wealth and desires a particular consumption bundle. We denote this trader with "N" and call him a noise trader. The noise trader demands the following consumption bundle:

$$C_n = (1 - \delta) \min(\delta, D), \quad 0 < \delta < 1$$

which is plotted in Figure 1-1(b). Figure 1-1(c) plots C_n as a fraction of the total consumption D . Since $C_n \leq \delta(1 - \delta)$, the wealth the noise trader needs to acquire the consumption bundle, is

$$W_n = E[C_n \cdot \phi] \leq E[\delta(1 - \delta) \cdot \phi] = \delta(1 - \delta) < \delta$$

where we have used the fact that the bond price is 1. The consumption for the representative agent (excluding the noise trade) is then $C = D - C_n$, also shown in Figure 1-1(b). The state price density in this case is

$$\phi = au'(C) = \frac{a}{D - (1 - \delta) \min(\delta, D)}$$

Since the price of the bond is one, we have

$$B = E[1 \cdot \phi] = \int_0^\delta \frac{a}{\delta D} (2D) dD + \int_\delta^1 \frac{a}{D - \delta(1 - \delta)} (2D) dD = 1$$

which gives

$$a = \frac{1}{4} \left\{ 1 - \frac{\delta}{2} + \frac{1}{2} \delta(1 - \delta) [\ln(1 - \delta + \delta^2) - 2 \ln(\delta)] \right\}^{-1}$$

As noted above, the wealth needed to acquire the consumption bundle C_n , W_n , is less than δ , so it is small if δ is small. The stock price in the presence of the noise trader is given by

$$\begin{aligned} S &= E[D \cdot \phi] = \int_0^\delta D \frac{a}{\delta D} (2D) dD + \int_\delta^1 D \frac{a}{(1 - \delta)\delta} (2D) dD \\ &= a \{ 1 + 3\delta - 5\delta^2 + 2\delta^3 + 2\delta^2(1 - \delta)^2 [\ln(1 - \delta + \delta^2) - 2 \ln(\delta)] \} = \frac{1}{4} + O(\delta) \end{aligned}$$

where $O(\delta)$ denotes terms of order δ or higher. Thus, $S/S^* = \frac{1}{2} + O(\delta)$. We can measure the impact of the noise trade on the stock price by

$$1 - \frac{S}{S^*} = \frac{1}{2} + O(\delta)$$

which remains non-negligible even when δ , and therefore the amount of wealth controlled by the noise trader, approaches zero.

This is a stark result: a price-taking trader with negligible wealth can exert finite influence on asset prices. The noise trader spends most of his wealth on consumption

in low-dividend states. Given that the marginal utility of the other traders in these states is very high, the state prices for these states are also high and, more importantly, a small change in the consumption level can change the state prices significantly. As we show above, the wealth required for the noise traders to finance their desired consumption profile is small, even though most of their consumption occurs in states with relatively high state prices.

While the above example is rather simple, its intuition holds more generally. In the case of logarithmic preferences, the state price density is proportional to the rational trader's marginal utility $u'(C)$: $\phi = au'(C)$, where a is the proportionality constant. When the irrational trader is introduced into the economy and he purchases ε units of state-contingent claims that pay off only when the aggregate consumption is C , the state-price density will change by $\Delta\phi \approx -au''(C)\varepsilon$. The total cost for the purchase is $w \equiv \phi\varepsilon \approx au'(C)\varepsilon$ when ε is small. Divided by the wealth spent by the irrational trader, we obtain the marginal change in the state-price density:

$$\frac{\Delta\phi}{w} = \frac{u''(C)}{u'(C)} = \frac{1}{C}$$

which is independent of ε . Clearly, in “bad” states, in which C is low (close to zero), irrational traders can have a large impact on the state-price density with little wealth if they decide to bet on these states. Through their impact on the state-price density in bad states, irrational traders can influence asset prices, such as the prices of the stock and the bond. Given that the bond is used as a numeraire and its price is always one, this influence is captured in the stock price, given by $S = E[D \cdot \phi]$, as shown above.

Our example clearly demonstrates the possibility of influencing asset prices with little wealth. The remaining question is whether such a situation can arise in “realistic” settings. In particular, for our purpose in this paper, can the irrational traders with incorrect beliefs maintain a significant price impact even as their relative wealth diminishes from investment losses in the market? In the remainder of the paper, we use a canonical model to address these questions.

1.3 The Model

We consider a standard setting similar to that of Black and Scholes (1973). For simplicity, we make the model parsimonious.

Information structure

The economy has a finite horizon and evolves in continuous time. Uncertainty is described by a one-dimensional, standard Brownian motion B_t for $0 \leq t \leq T$, defined on a complete probability space (Ω, F, P) , where F is the augmented filtration generated by B_t .

The financial market

There is a single share of a risky asset in the economy, the stock, which pays a terminal dividend payment D_T at time T , determined by process:

$$dD_t = D_t(\mu dt + \sigma dB_t) \tag{1.1}$$

where $D_0 = 1$ and $\sigma > 0$. There is also a zero coupon bond available in zero net supply. Each unit of the bond makes a sure payment of one at time T . We use the risk-free bond as the numeraire and denote the price of the stock at time t by S_t .

Endowments

There are two competitive traders in the economy, each endowed with a half share of the stock (and none of the bond) at time zero.

Trading strategies

The financial market is frictionless and has no constraints on lending and borrowing. Traders' trading strategies satisfy the standard integrability condition used to avoid pathologies:

$$\int_0^T \theta_t^2 d\langle S \rangle_t < \infty$$

where θ_t is the number of stock shares held in the portfolio at time t and $\langle S \rangle_t$ is the quadratic variation process of S_t (see, e.g., Duffie and Huang (1986) and Harrison and Kreps (1979)).

Preferences and beliefs

Both traders have constant relative risk aversion utility over their consumption at time T :

$$\frac{1}{1-\gamma} C_T^{1-\gamma}, \quad \gamma \geq 1.$$

For ease of exposition, we only consider the cases when $\gamma \geq 1$. The cases when $0 < \gamma < 1$ can be analyzed similarly and the results are similar in spirit.

Standard aggregation results imply that each trader in our model can actually represent a collection of traders with the same preferences. This provides a justification for our competitive assumption for each of the traders. The first trader, the rational trader, knows the true probability measure P and maximizes expected utility

$$E_0^P \left[\frac{1}{1-\gamma} C_{r,T}^{1-\gamma} \right] \tag{1.2}$$

where the subscript r denotes quantities associated with the rational trader. The second trader, the irrational trader, believes incorrectly that the probability measure is Q , under which

$$dB_t = (\sigma\eta)dt + dB_t^Q \tag{1.3}$$

and hence

$$dD_t = D_t \left[(\mu + \sigma^2\eta) dt + \sigma dB_t^Q \right] \tag{1.4}$$

where B_t^Q is the standard Brownian motion under the measure Q and η is a constant, parameterizing the degree of irrationality of the irrational trader. When η is positive, the irrational trader is optimistic about the prospects of the economy and overestimates the rate of growth of the aggregate endowment. Conversely, a negative η corresponds to a pessimistic irrational trader. The irrational trader maximizes

expected utility using belief Q :

$$\mathbb{E}_0^Q \left[\frac{1}{1-\gamma} C_{n,T}^{1-\gamma} \right] \quad (1.5)$$

where the subscript n denotes quantities associated with the irrational trader.

Because η is assumed to be constant, the probability measure of the irrational trader Q is absolutely continuous with respect to the objective measure P , i.e., both traders agree on zero-probability events. Let $\xi_t \equiv (dQ/dP)_t$ denote the density (Radon-Nikodym derivative) of the probability measure Q with respect to P ,

$$\xi_t = e^{-\frac{1}{2}\eta^2\sigma^2t + \eta\sigma B_t}. \quad (1.6)$$

The irrational trader maximizes

$$\mathbb{E}_0^Q \left[\frac{1}{1-\gamma} C_{n,T}^{1-\gamma} \right] = \mathbb{E}_0^P \left[\xi_T \frac{1}{1-\gamma} C_{n,T}^{1-\gamma} \right]. \quad (1.7)$$

This permits us to interpret the objective of the irrational trader as the expected value of a state-dependent utility function, $\xi_T \frac{1}{1-\gamma} C_{n,T}^{1-\gamma}$, under the true probability measure P .

The equivalence between incorrect beliefs and state dependent preferences raises a conceptual question about the precise definition of irrationality. It is beyond the scope of this paper to address this question, and our analysis of this form of irrationality is primarily motivated by the fact that it is widely adopted in the recent literature on behavioral models of asset prices.

1.4 The Equilibrium

The competitive equilibrium of the economy defined above can be solved analytically. Since there is only one source of uncertainty in the economy, the financial market is dynamically complete as long as the volatility of stock returns remains non-zero

almost surely. Consequently, the equilibrium allocation is efficient and can be characterized as the solution to a central planner's problem:

$$\max \left[\frac{1}{1-\gamma} C_{r,T}^{1-\gamma} + b \xi_T \frac{1}{1-\gamma} C_{n,T}^{1-\gamma} \right] \quad (1.8a)$$

$$\text{s.t. } C_{r,T} + C_{n,T} = D_T \quad (1.8b)$$

where b is the ratio of the utility weights for the two traders. The equilibrium allocation is characterized in the following proposition.

Proposition 1 *For the economy defined in Section 1.3, the equilibrium allocation between the two traders is*

$$C_{r,T} = \frac{1}{1 + (b \xi_T)^{1/\gamma}} D_T \quad (1.9a)$$

$$C_{n,T} = \frac{(b \xi_T)^{1/\gamma}}{1 + (b \xi_T)^{1/\gamma}} D_T \quad (1.9b)$$

where

$$b = e^{(\gamma-1)\eta\sigma^2 T}. \quad (1.10)$$

The price of a financial security with the terminal payoff Z_T is given by

$$P_t = \frac{\mathbb{E}_t \left[\left(1 + (b \xi_T)^{1/\gamma} \right)^\gamma D_T^{-\gamma} Z_T \right]}{\mathbb{E}_t \left[\left(1 + (b \xi_T)^{1/\gamma} \right)^\gamma D_T^{-\gamma} \right]}. \quad (1.11)$$

For the stock, $Z_T = D_T$ and its return volatility is bounded between σ and $\sigma(1 + |\eta|)$.

Since the instantaneous volatility of stock returns is bounded below by σ , the stock and the bond dynamically complete the financial market. In the limiting cases when only the rational or the irrational trader is present, the stock prices, denoted

by S_t^* and S_t^{**} , respectively, are given by

$$S_t^* = e^{(\mu/\sigma^2 - \gamma)\sigma^2 T + \frac{1}{2}(2\gamma - 1)\sigma^2 t + \sigma B_t} \quad (1.12a)$$

$$S_t^{**} = e^{(\mu/\sigma^2 - \gamma + \eta)\sigma^2 T + \frac{1}{2}[(2\gamma - 1) - 2\eta]\sigma^2 t + \sigma B_t} = S_t^* e^{\eta\sigma^2(T-t)}. \quad (1.12b)$$

We will use this equilibrium model to analyze the survival and extinction of the traders. We employ the following common definition of extinction, and, conversely, of survival.

Definition 1 *The irrational trader is said to experience relative extinction in the long-run if*

$$\lim_{T \rightarrow \infty} \frac{C_{n,T}}{C_{r,T}} = 0 \quad a.s. \quad (1.13)$$

The relative extinction of the rational trader can be defined symmetrically. A trader is said to survive relatively in the long-run if relative extinction does not occur.

In the above definition and throughout the paper, all limits are understood to be almost sure (under the true probability measure P) unless specifically stated otherwise.

In our model, the final wealth of each trader equals their terminal consumption. Thus, the definition of survival and extinction is equivalent to a similar definition in terms of wealth.

1.5 Logarithmic Preferences

We first consider the case where both the rational and the irrational traders have logarithmic preferences. We have the following result:

Proposition 2 *Suppose $\eta \neq 0$. For $\gamma = 1$, the irrational trader never survives.*

This result is immediate. For $\gamma = 1$, the rational trader holds the portfolio with maximum expected growth (see, e.g., Hakansson (1971)). Any deviation in beliefs

from the true probability causes the irrational trader to move away from the maximum growth portfolio, which leads to his long-run relative extinction.

Our interest here, however, is not on the survival of the irrational trader, but on the impact of irrationality on the long-run stock price. Under logarithmic preferences, $b = 1$ and from Proposition 2 the stock price is

$$S_t = \frac{1 + \xi_t}{\mathbb{E}_t [(1 + \xi_T)/D_T]} = \frac{1 + \xi_t}{1 + e^{-\eta\sigma^2(T-t)}\xi_t} S_t^* \quad (1.14)$$

where S_t^* denotes the stock price in an identical economy populated only by the rational trader, given in (1.12). We now prove that the irrational trader can maintain a large impact on the stock price despite losing most of his wealth. To state our result formally, we define the relative wealth shares of the rational and irrational traders, respectively,

$$\alpha_{n,t} \equiv \frac{W_{n,t}}{W_{r,t} + W_{n,t}} = \frac{\xi_t}{1 + \xi_t}, \quad \alpha_{r,t} \equiv 1 - \alpha_{n,t}$$

The price impact the irrational trader can be measured by $1 - \frac{S_t}{S_t^*}$, the relative deviation in stock price from its limiting value with only the rational trader. We have

Proposition 3 *Consider the case of a pessimistic irrational trader, $\eta < 0$. For any ε as small as $e^{-\frac{\sigma^2\eta^2}{12(1+|\eta|)}T}$, there exists a point in time $t \geq T/(1 + |\eta|)$, such that*

$$\text{Prob} [\alpha_{n,t} \geq \varepsilon] \leq \varepsilon \quad (1.15a)$$

$$\text{Prob} \left[1 - \frac{S_t}{S_t^*} \leq 1 - \varepsilon \right] \leq \varepsilon. \quad (1.15b)$$

Intuitively, Proposition 3 shows that after a long period of time, which constitutes a nontrivial fraction of the horizon of the economy, the relative wealth of the irrational trader is most likely to be very small (which is consistent with his long-run extinction), but his impact on the stock price is most likely to remain large (the ratio S_t/S_t^* stays far away from one).

Another way to illustrate the persistent nature of the irrational trader's price impact is by examining the long-run behavior of the instantaneous moments of stock

returns, which can be derived explicitly. For example, the conditional volatility of stock returns is

$$\sigma_{S,t} = \sigma + \eta\sigma\alpha_{n,t} - \eta\sigma \left[1 - \frac{1}{1 + e^{-\eta\sigma^2(T-t)}\alpha_{n,t}(1 - \alpha_{n,t})^{-1}} \right]$$

and the conditional mean is

$$\mu_{S,t} = \sigma_{S,t}^2 - \alpha_{n,t}\eta\sigma\sigma_{S,t}.$$

To visualize the behavior of stock return moments, consider the following numerical example. The irrational trader is assumed to be pessimistic ($\eta = -2$). The horizon of the economy is set to $T = 400$, so the relative wealth of the irrational trader becomes relatively small long before the final date. We let the current time t be sufficiently large, so with high probability most of wealth in the economy is controlled by the rational trader. For convenience, we define the following normalized state variable:

$$g_{s,t} \equiv \frac{B_t - B_s}{\sqrt{t - s}} \quad (1.16)$$

where $s < t$. It is easy to show that $g_{s,t}$ is the unanticipated dividend growth normalized by its standard deviation, which has a standard normal distribution. Figure 1-2 plots the Sharpe ratio of instantaneous stock returns and the wealth distribution between the two traders at $t = 150$ against the normalized state variable $g_{0,t}$. The probability density for $g_{0,t}$ is illustrated by the shaded area (with the vertical axis on the right). The bottom panel of Figure 1-2 shows that with almost probability one, the wealth of the economy is all controlled by the rational trader at this time. Yet as the top panel of the figure shows, the conditional Sharpe ratio of stock returns is very different from σ , which is its value in the economy populated only by the rational trader. In particular, over a large range of values of the dividends, the conditional Sharpe ratio of returns is approximately equal to $\sigma(1 - \eta) \neq \sigma$.

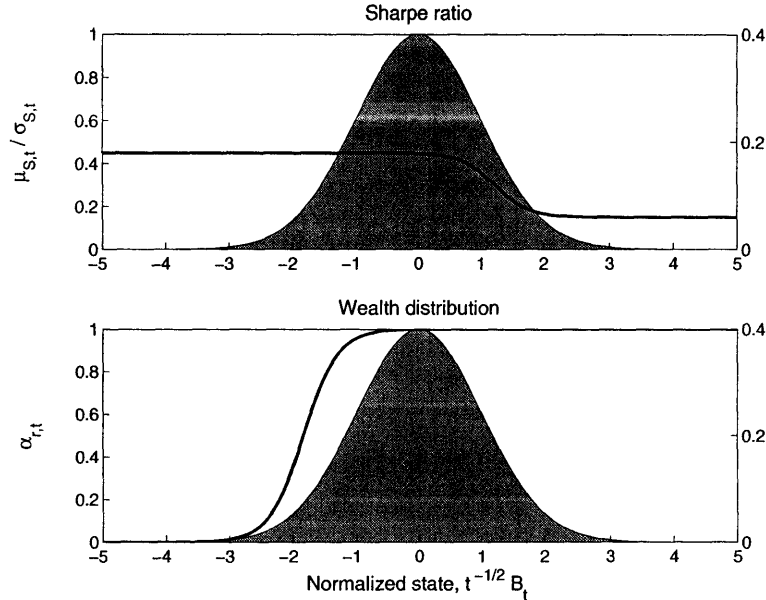


Figure 1-2: The conditional Sharpe ratio of stock returns, $\mu_{S,t}/\sigma_{S,t}$ and the wealth distribution $\alpha_{r,t} = W_{r,t}/(W_{r,t} + W_{n,t})$ are plotted against the normalized state variable, $g_{0,t} \equiv B_t/\sqrt{t}$. The shaded area is the probability density function of the normalized state variable (vertical axis on the right). The model parameters are set at $\mu = 0.05$, $\sigma = 0.15$, $\eta = -2$, $T = 400$ and $\gamma = 1$. The current time is $t = 150$.

Figure 1-3 provides a complimentary illustration. It shows the most likely path over time (the path with highest probability) for the irrational trader's wealth share and the Sharpe ratio of stock returns. In fact, the irrational trader's wealth share diminishes to zero exponentially while his price impact diminishes at a much slower rate. The Sharpe ratio stays away from its level in an economy without an irrational trader for an extended period of time before eventually converging to the limiting value.

In order to better understand how the irrational trader can exert influence on the stock price despite having negligible wealth, we examine how his presence affects the state price density (SPD). The left panels of Figure 1-4 plot the relative consumption shares of the rational and the irrational traders at two different times, $t = 0, 25$, as a function of the normalized state variable, $g_{t,T}$, i.e., the normalized unanticipated dividend growth from t to T defined in (1.16). At each date, the state of the economy is conditioned on $B_t = 0$, the most likely state. For $t = 0$, the irrational trader owns

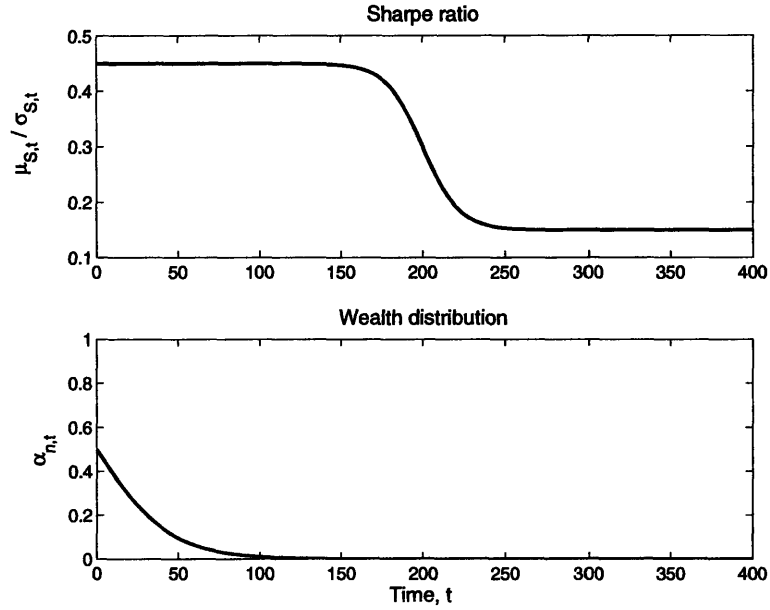


Figure 1-3: The maximum likelihood path of the irrational trader's wealth share, $\alpha_{n,t} = W_{n,t}/(W_{r,t} + W_{n,t})$, and the Sharpe ratio, $\mu_{S,t}/\sigma_{S,t}$. The model parameters are set at $\mu = 0.05$, $\sigma = 0.15$, $\eta = -2$, $T = 400$ and $\gamma = 1$.

half of the economy. But at $\eta = -4$, he is very pessimistic and bets on states of low dividends (states toward the left end of the horizontal axis). This is shown in the top left panel of Figure 1-4. The dashed line plots his terminal consumption for different states of the economy. It is worth pointing out that the consumption choice of the irrational trader in this economy is similar to that in the simple one-period economy we considered in Section 1.2, as shown in Figure 1-1(c), where the irrational trader consumes a share of $1 - \delta$ of the aggregate endowment in states with low dividends and much smaller share in other states. This explains why in both economies the irrational trader can exert significant influence on prices despite of being left with relatively little wealth.

Over time, the 'bad' states become less likely and the irrational trader's bets become less valuable. Thus, his wealth decreases. At $t = 25$ and $B_t = 0$, these bad states become extremely unlikely and the irrational trader has lost most of his wealth. His wealth as fraction of total wealth has fallen from $1/2$ at $t = 0$ to 0.01 . As shown in the bottom left panel of Figure 1-4, going forward, irrational trader consumes a

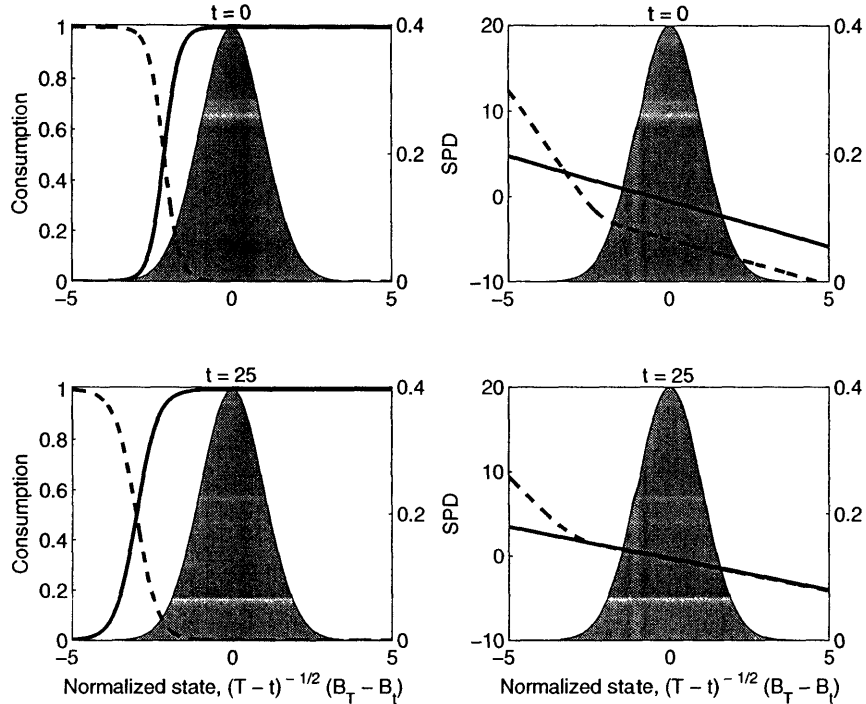


Figure 1-4: The terminal consumption of the rational and irrational traders as a fraction of the total consumption and the state price density (SPD) in different terminal states of the economy at different times. The model parameters are set to be $\mu = 0.05$, $\sigma = 0.15$, $\eta = -4$, $\gamma = 1$ and $T = 50$. The horizontal axis in all panels is the normalized state variable $g_{t,T} \equiv (B_T - B_t)/\sqrt{T-t}$, which has a standard normal distribution with zero mean and unit variance, which is shown by the shaded area (vertical axis on the right). In the two panels on the left, the terminal consumption for the rational trader (the solid line) and the irrational trader (the dotted line) are plotted against the normalized state variable at times $t = 0, 25$, respectively, when $B_t = 0$. In the two panels on the right, the dashed line plots the logarithm of the state price density at times $t = 0, 25$, respectively, which is $\ln \{[(1 + \xi_T)/D_T]/E_t [(1 + \xi_T)/D_T]\}$. The solid line plots the logarithm of the state price density in the economy populated only by the rational traders, which is $\ln \{D_T^{-1}/E_t [D_T^{-1}]\}$.

non-trivial fraction of the total wealth only in the extreme states toward the left end of the horizontal axis. The probability of these states, as shown by the shaded area, becomes very small and so is the irrational trader's wealth.

In the two panels on the right of Figure 1-4, we plot the state price density against the normalized state variable $g_{t,T}$ at the two times, $t = 0, 25$, conditioned again on $B_t = 0$. With logarithmic preferences, the equilibrium state price density at time t is given by

$$\phi_t \equiv \frac{(1 + \xi_T)D_T^{-1}}{E_t [(1 + \xi_T)D_T^{-1}]}$$

which is represented by the dashed line in each of the two panels. The solid line plots the state price density when the economy is populated only by the rational traders, which can be obtained by setting $\xi_T = 0$ in the above expression for ϕ_t . The top panel gives the state price density at $t = 0$. At this point, the irrational trader has a half share of the total wealth and his portfolio policy has a significant influence on the state price density over the whole range. In particular, being pessimistic, he is effectively betting on the bad states, which causes the state price density to increase for the bad states and decrease for the good states. This is shown by the difference between the dashed line, the state price density in the presence of the irrational trader, and the solid line, the state price density without the irrational trader. As time passes, the irrational trader's wealth dwindles and his influence on the state price density diminishes quickly for most of the states, as the bottom panel for $t = 25$ shows. However, for the extremely bad states his influence remains significant because he is still betting heavily on these states.

We can show that the price impact of the irrational trader with negligible wealth does not rely on excessive leverage. The fraction of irrational trader's wealth invested in the stock is given by $\sigma_{S,t} + \eta\sigma(1 - \alpha_{n,t})$, which is bounded in absolute value by $\sigma(1 + 2|\eta|)$. The irrational trader can make bets on states with low aggregate endowment not by taking extreme portfolio positions, but rather by under-weighting the stock in his portfolio over long periods of time.

The simple case of logarithmic preferences developed above clearly shows that survival and price impact are in general not equivalent. In particular, survival is not a necessary condition for the irrational trader to influence long-run prices, and depending on their beliefs, irrational traders can maintain a significant price impact even as their wealth becomes negligible over time.

In the remaining sections, we consider the general case when $\gamma \neq 1$ and analyze the survival of the irrational trader, his price impact, and his portfolio choices.

1.6 Survival

In the case of logarithmic preferences, the irrational trader does not survive in the long-run simply because his portfolio grows more slowly than the maximum growth rate, the rate achieved by the rational trader. For the coefficient of relative risk aversion different from one, though, the rational trader no longer holds the optimal growth portfolio and under an incorrect belief, the irrational trader may end up holding a portfolio that is closer to the optimal growth portfolio, and so his wealth may grow more rapidly. This was the argument put forward by DSSW using a partial equilibrium setting. In this section, we examine the long-run survival of the irrational trader in our general equilibrium setting.

From the competitive equilibrium derived in Section 1.4, we have the following result:

Proposition 4 *Suppose $\eta \neq 0$. Let $\eta^* = 2(\gamma - 1)$. For $\gamma > 1$ and $\eta \neq \eta^*$, only one of the traders survives in the long run. In particular, we have*

$$\begin{aligned}
 \text{Pessimistic irrational trader: } \eta < 0 & \Rightarrow \text{Rational trader survives} \\
 \text{Moderately optimistic irrational trader: } 0 < \eta < \eta^* & \Rightarrow \text{Irrational trader survives} \\
 \text{Strongly optimistic irrational trader: } \eta > \eta^* & \Rightarrow \text{Rational trader survives.}
 \end{aligned}
 \tag{1.17}$$

For $\eta = \eta^$, both rational and irrational traders survive.*

For $\gamma > 1$, Proposition 4 identifies three distinct regions in the parameter space as shown in Figure 1-5. For $\eta < 0$, the irrational trader is pessimistic and does not survive in the long-run. For $0 < \eta < \eta^*$, the irrational trader is moderately optimistic and survives in the long-run while the rational trader does not. For $\eta > \eta^*$, the irrational trader is strongly optimistic and does not survive. Clearly, other than the knife-edge case ($\eta = \eta^*$), only one of the traders can survive.

In order to gain more insight into what determines the survival of each type of traders, we examine their terminal wealth (consumption) profiles. The two panels on

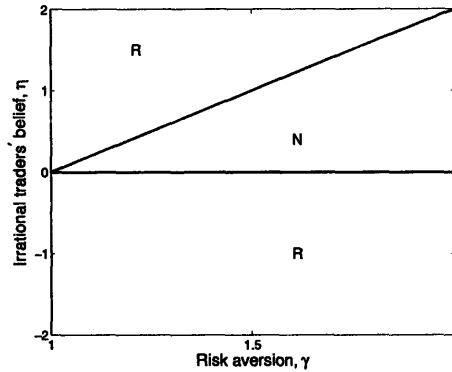


Figure 1-5: The survival of rational and irrational traders for different values of η and γ . For each region in the parameter space, we document which of the agents survives in the long run. “R” means that survival of the rational trader is guaranteed inside the region, “N” corresponds to the irrational trader.

the left in Figure 1-6 show the two traders’ terminal wealth profiles for two values of T (10 and 30) when the irrational trader is pessimistic. The solid line shows the terminal wealth share of the rational trader and the dashed line shows that for the irrational trader. As expected, the rational trader ends up with more wealth in good states of the economy (when the dividend is high) while the irrational trader, being pessimistic, ends up with more wealth in the bad states of the economy. As the horizon increases, the irrational trader ends up with non-trivial wealth in more extreme and less likely, low dividend states. When the irrational trader is mildly optimistic, the situation is different. His impact on the prices makes the bad states (i.e., the low dividend states) cheaper than the good states. This induces the rational trader to accumulate more wealth in the bad states by giving up wealth in the good states, including those with high probabilities. As a result, the irrational trader is more likely to end up with more wealth. When strongly optimistic, the irrational trader ends up accumulating wealth in very unlikely, good states by giving up wealth in most other states, which leads to his extinction in the long-run.

It is important to recognize that our results on the long-run survival of irrational traders are obtained in absence of intermediate consumption. In other words, these results are primarily driven by the portfolio choices of different traders in the market and their impact on prices. This allows us to focus on how irrational beliefs influence

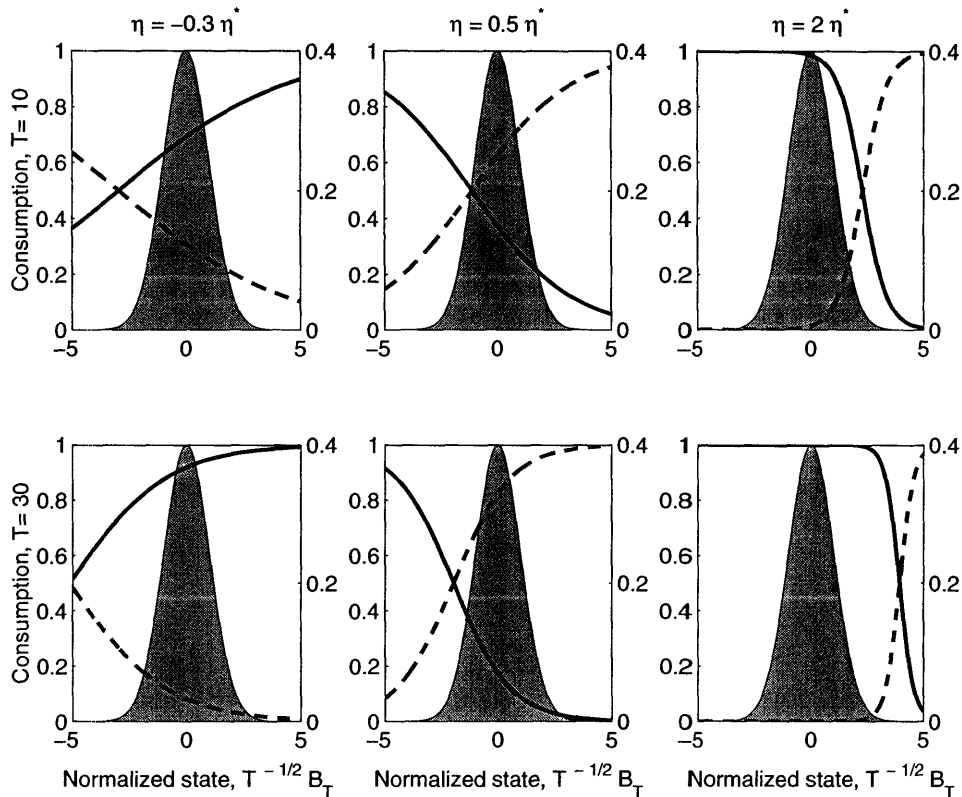


Figure 1-6: The terminal consumption of rational and irrational traders for different horizon T . We consider two values of T , 10 and 30, respectively. The model parameters are set at $\mu = 0.12$, $\sigma = 0.18$, and $\gamma = 5$. We consider three distinctive cases for the irrational trader's belief: (1) pessimistic, $\eta = -0.3\eta^*$, (2) moderately optimistic, $\eta = 0.5\eta^*$, and (3) strongly optimistic, $\eta = 2\eta^*$. The horizontal axis in all panels is the normalized value of the terminal dividend, i.e., $[\ln D_T - (\mu - \frac{1}{2}\sigma^2)T] / (\sigma\sqrt{T})$, which has a standard normal distribution with zero mean and unit variance, shown by the shaded area (vertical axis on the right). The two panels on the left show the terminal consumption, as a fraction of the total consumption, of the rational trader (solid line) and the irrational trader (dashed line) with a pessimistic belief, i.e., $C_{r,T}/D_T$ and $C_{n,T}/D_T$, for the two values of the horizon, $T = 10, 30$, respectively. The two panels in the middle and on the right show the terminal consumption, as a fraction of the total consumption, of the rational trader and the irrational trader with a moderately and strongly optimistic beliefs for the two values of T , respectively.

traders' trading behavior and how it along affects their wealth evolution. When intermediate consumption is allowed, traders' consumption policies will also be affected by their beliefs, which can significantly affect their wealth accumulation as well. The net impact of irrational belief on a trader's wealth evolution depends on how it affects his portfolio choice *and* his consumption choice. Using an infinite horizon setting with intermediate consumption, Blume and Easley (2001) and Sandroni (2000) have shown that traders with (persistently) irrational beliefs will not survive while traders with rational beliefs will. Their analysis clearly shows that the influence of incorrect beliefs on the irrational traders' consumption policy can reduce their chance of survival. However, their result critically relies on several conditions imposed on the traders' preferences and aggregate endowments. For example, they require that aggregate endowment is bounded above and below, away from zero. When these bounds are not imposed, as is the case in this paper, traders with rational beliefs may not always survive while traders with irrational beliefs may.² To provide a comprehensive analysis of the survival conditions with intermediate consumption is beyond the scope of this paper and is left for future research. But it suffices to say that even with intermediate consumption, the long-run survival of irrational traders is possible in absence of further restrictions on preferences and/or endowments.

Another difference between our setting and that of Blume and Easley (2001) is that we use a particular and simple form of beliefs of the irrational traders. In our model, such traders maintain a constant belief about the drift of the endowment process and do not update their belief based on the realized data. To maintain analytical tractability, we do not allow for a more general form of beliefs, e.g., that resulting from inefficient learning. However, in the setting of Blume and Easley (2001), the specific form of the belief process is less important for the survival results than the aggregate endowment process and agents' preferences. Based on this observation, we would expect the intuition of our model to apply to more general settings as well, in

²In a simple case considered by Wang (1996), even among rational traders, survival depends on preferences. In our setting, we did not impose any upper or positive lower bounds on endowments.

particular to certain types of inefficient learning.

1.7 The Price Impact of Irrational Traders

We have already seen in the case of logarithmic preferences that the irrational trader's influence on prices does not decay as quickly as his relative wealth share. In this section, we extend our analysis to the general case for γ and characterize the precise combinations of model parameters under which such phenomenon is possible.

Our interest is in the behavior of prices in the long run when the horizon of the economy, T , is long. In order to obtain an explicit characterization, we look at the limit when T approaches infinity and derive from the limit an analytical approximation for a large, but finite T . By the definition of the limit, this approximation becomes arbitrarily accurate when T is sufficiently large. Specifically, we call two stochastic processes asymptotically equivalent if for large values of T , their ratio converges to unity with probability one.

Definition 2 *Two stochastic processes, X_t and Y_t , are asymptotically equivalent if*

$$\lim_{T \rightarrow \infty} \frac{X_T}{Y_T} = 1 \quad a.s.$$

which we denote $X_T \sim Y_T$.

When studying an economy with a long horizon, T , we need to have a sense about what it means for a particular property of the economy to persist for a significant period of time. Suppose, for example, we claim that the irrational trader's influence on a variable is significant as long as the variable exceeds a fixed level e within a time interval. Such an influence is persistent only if for a larger T , the corresponding time interval of the irrational trader's influence also increases in proportion. Otherwise, the fraction of time the irrational trader does have an influence becomes smaller for a larger T and thus his influence is only transitory and negligible.

To make this more formal, we consider the current time of observation to be $t = \lambda T$, $0 < \lambda \leq 1$. As T grows, the “current” time t increases as well, but it remains at a constant fraction of the horizon of the economy. Moreover, the time remaining until the final date of the economy is also increasing proportionally to T . Since the properties of the equilibrium prices and quantities depend on how much time is remaining until the final date, they depend on λ .

We define three values of λ to help us characterize points of change in the limiting behavior:

$$\lambda_S \equiv \frac{2}{2\gamma - \eta}, \quad \lambda_r \equiv \frac{\eta}{(\gamma - 1)(2\gamma - \eta)}, \quad \lambda_n \equiv \frac{\eta}{\eta(\gamma + 1) - 2\gamma(\gamma - 1)}. \quad (1.18)$$

It is easy to verify that for $\eta < \eta^*$, $0 < \lambda_S \leq 1$; for $0 < \eta \leq \eta^*$, $0 < \lambda_r \leq 1$; and for $\eta < 0$ or $\eta > \eta^*$, $0 < \lambda_n \leq 1$. The limiting behavior of the stock price process can be characterized as follows.

Proposition 5 *At $t = \lambda T$, the stock price behaves as follows:*

Case 1. Pessimistic Irrational Trader ($\eta < 0$):

$$S_t \sim \begin{cases} S_t^* e^{\eta[\sigma^2 T + \frac{1}{2}(\eta - 2\gamma)\sigma^2 t - \sigma B_t]}, & 0 < \lambda < \lambda_S \\ S_t^*, & \lambda_S < \lambda \leq 1 \end{cases}$$

Case 2. Moderately Optimistic Irrational Trader ($0 < \eta < \eta^$):*

$$S_t \sim \begin{cases} S_t^* e^{\eta[(\gamma - 1 - \frac{1}{2}\eta)\sigma^2 t + \sigma B_t]}, & 0 < \lambda < \lambda_S \\ S_t^{**}, & \lambda_S < \lambda \leq 1 \end{cases}$$

Case 3. Strongly Optimistic Irrational Trader ($\eta^ < \eta$):*

$$S_t \sim S_t^*.$$

The values of the stock price in homogeneous economies, S_t^ and S_t^{**} , are given in*

Equation (1.12). The asymptotic values of the instantaneous moments of stock returns are equal to the moments of the corresponding asymptotic expressions for stock prices above.

Observe that in the first two cases, when the irrational trader is pessimistic or moderately optimistic, the stock price process does not converge quickly to its value in the economy populated exclusively by the rational trader who survives in the long-run. Instead, over long periods of time, i.e., for t between 0 and $\lambda_S T$, the stock price process is affected by the presence of both traders. This can occur even when the wealth of the irrational trader becomes negligible way before $\lambda_S T$.³ We thus have generalized the results obtained in the context of a log-utility economy. A trader can control an asymptotically infinitesimal fraction of the total wealth and yet exert a non-negligible effect on the stock price. In other words, convergence in wealth does not readily imply convergence in prices.

1.8 Portfolio Policies

Proposition 5 in the previous section established the possibility that a trader whose wealth diminishes over time can have a persistent impact on asset prices. In this section, we study the traders' portfolio policies. In particular, we show that convergence in the price process does not lead to immediate convergence in policies, which is another and somewhat subtle channel through which traders with asymptotically infinitesimal wealth may affect the long-run behavior of the economy. Moreover, by characterizing the portfolio policy one gains an alternative view on long-run survival in equilibrium, which is complementary to the analysis of state-contingent consumption choices in sections 1.5 and 1.6.

Expressions for portfolio policies are not available in closed form. However, us-

³ For brevity, we have omitted the discussion of wealth distribution over time. Interested readers can refer to our working paper, Kogan, Ross, Wang and Westerfield (2003), where we show that for cases 1 and 3, the irrational trader's wealth is asymptotically negligible for any time λT with $\lambda < \lambda_S$.

ing the similar argument as in the proof of the bound on stock price volatility in Proposition 1, we can establish the following result:

Proposition 6 *For both traders, their portfolio weight in the stock, denoted by w , is bounded:*

$$|w| \leq 1 + |\eta|(\gamma + 1)/\gamma. \quad (1.19)$$

The bound on the traders' portfolio holdings is important for our results. It explicitly shows that price impact of the irrational trader with negligible wealth does not rely on excessive leverage. It also implies that our long-run survival results do not rely on the traders' use of high leverage. Our solution for the equilibrium remains valid even if traders are constrained in their portfolio choices, as long as the constraint is sufficiently loose to allow for $w = \pm[1 + |\eta|(\gamma + 1)/\gamma]$.

To analyze the traders' portfolio policies in more detail, we decompose a trader's stock demand into two components, the myopic component and the hedging component. The sum of the two gives the trader's total stock demand. We have the following proposition.

Proposition 7 *At $t = \lambda T$, the individual stock holdings behave as follows:*

Case 1. Pessimistic Irrational Trader ($\eta < 0$):⁴

$$w_{r,t} \sim \begin{cases} \begin{array}{ccc} \text{(myopic)} & \text{(hedging)} & \text{(total)} \\ \frac{\gamma-\eta}{\gamma(1-\eta)} & - \frac{(\gamma-1)\eta}{\gamma(1-\eta)} & = 1, & 0 < \lambda < \lambda_S \\ 1 & + 0 & = 1, & \lambda_S < \lambda \leq 1 \end{array} \\ \\ \begin{array}{ccc} \text{(myopic)} & \text{(hedging)} & \text{(total)} \\ \frac{1}{1-\eta} & + 0 & = \frac{1}{1-\eta}, & 0 < \lambda < \min(\lambda_n, \lambda_S) \\ 1 + \frac{\eta}{\gamma} & + 0 & = 1 + \frac{\eta}{\gamma}, & \max(\lambda_n, \lambda_S) < \lambda \leq 1 \end{array} \end{cases}$$

⁴The limit of the portfolio policy for values of $\lambda \in [\min(\lambda_n, \lambda_S), \max(\lambda_n, \lambda_S)]$ can be characterized explicitly as well, but the results depend on the ordering between λ_n and λ_S , which in turn is determined by the values of model parameters. We omit these results to simplify the exposition.

Case 2. Moderately Optimistic Irrational Trader ($0 < \eta < \eta^$):*

$$w_{r,t} \sim \begin{cases} \begin{array}{l} \text{(myopic)} \quad \text{(hedging)} \quad \text{(total)} \\ \frac{1}{1+\eta} + 0 = \frac{1}{1+\eta}, \quad 0 < \lambda < \lambda_r \\ \frac{1}{1+\eta} + \frac{\eta(\gamma-1)}{\gamma(1+\eta)} = 1 - \frac{\eta}{\gamma(1+\eta)}, \quad \lambda_r < \lambda < \lambda_S \\ 1 - \frac{\eta}{\gamma} + 0 = 1 - \frac{\eta}{\gamma}, \quad \lambda_S < \lambda \leq 1 \end{array} \\ \\ \begin{array}{l} \text{(myopic)} \quad \text{(hedging)} \quad \text{(total)} \\ \frac{\gamma+\eta}{\gamma(1+\eta)} + \frac{\eta(\gamma-1)}{\gamma(1+\eta)} = 1, \quad 0 < \lambda < \lambda_S \\ 1 + 0 = 1, \quad \lambda_S < \lambda \leq 1 \end{array} \end{cases}$$

Case 3. Strongly Optimistic Irrational Trader, ($\eta^ < \eta$):*

$$w_{r,t} \sim 1 + 0 = 1, \quad 0 < \lambda \leq 1$$

$$w_{n,t} \sim \begin{cases} \begin{array}{l} \text{(myopic)} \quad \text{(hedging)} \quad \text{(total)} \\ 1 + \frac{\eta}{\gamma} + \frac{\eta(\gamma-1)}{\gamma} = 1 + \eta, \quad 0 < \lambda < \lambda_n \\ 1 + \frac{\eta}{\gamma} + 0 = 1 + \frac{\eta}{\gamma}, \quad \lambda_n < \lambda \leq 1 \end{array} \end{cases}$$

Since the moments of stock returns are asymptotically state-independent, it is intuitive to expect that the implied portfolio policies are myopic. Proposition 7 shows, however, that this is not true. In other words, the asymptotic portfolio policy can differ significantly from what the asymptotic moments of stock returns suggest. Such a surprising behavior can only be due to the hedging component of the traders' portfolio holdings since, by definition, the myopic component of portfolio holdings depends only on the conditional mean and variance of stock returns. Given that the instantaneous moments of stock returns are asymptotically state-independent, it may seem surprising that the hedging component of portfolio holdings remain finite, as

Case 3 in Proposition 7 illustrates for the irrational trader. The reason behind this result is that instantaneous moments of stock returns do not fully characterize the investment opportunities traders face. In particular, moments of stock returns do not always stay constant. As we have seen in Figure 1-2, for example, return volatility can change significantly as the relative wealth distribution changes. After a long time, the likelihood of the reversal of wealth distribution between the rational and irrational traders and a shift in return moments is relatively low. Nonetheless, the possibility of such a change remains important, which gives rise to the significant hedging demand in the traders' portfolio holdings.

Figure 1-7 illustrates the behavior of the economy when the irrational trader is strongly optimistic ($\eta > \eta^*$). In this case (Case 3 in Propositions 4, 5 and 7), the irrational trader does not survive and has no price impact in the long-run. For the chosen set of parameter values, $\lambda_n = 0.29$. The time of observation t is set to be $0.15T$. Thus $t < \lambda_n T$. As the bottom panel of Figure 1-7 shows, with almost probability one, the rational trader controls most of the wealth in the economy by this point in time. From Proposition 5, at this point the stock price converge closely to the price in the economy populated by only the rational trader. If we consider the Sharpe ratio of the stock, defined by μ_S/σ_S , which characterizes the instantaneous investment opportunity traders face, it also converges to its value of $\gamma\sigma$ in the limiting economy with the rational trader only. The top panel of Figure 1-7 plots the value of the Sharpe ratio for different states of the economy at time t . It is obvious that with almost probability one, the value of the Sharpe ratio equals its limit $\gamma\sigma$ (the probability distribution of the state of economy is shown by the shaded area). However, for very large values of D_t (or B_t), the economy will be dominated by the irrational trader (as we see from the bottom panel) and the instantaneous Sharpe ratio of the stock converges to its value in an economy populated by the irrational trader only, which is $(\gamma - \eta)\sigma$. Such a possibility, even though with very low probability under the true probability measure, can be important to the irrational trade because under his belief, its likelihood can be non-trivial. As a result, it can have a significant impact on the

irrational trader's portfolio choice.

The importance of these low probability but large changes in the Sharpe ratio is reflected in the traders' value function, which is given by

$$V(t, W_t, D_t) \equiv \mathbb{E}_t \left[\xi_T \frac{W_T^{1-\gamma}}{1-\gamma} \right] = \frac{1}{1-\gamma} e^{h(t, D_t)} W_t^{1-\gamma} \equiv \mathbb{E}_t \left[\xi_T \frac{1}{1-\gamma} C_{n,T}^{1-\gamma} \right]. \quad (1.20)$$

State dependence of the indirect utility function, i.e., the effect of possible changes in the Sharpe ratio, is captured by the function $h(t, D_t)$. The second panel of Figure 1-7 shows that for the irrational trader h is non-constant over a wide range of values of D_t . It exhibits significant state-dependence even when the contemporaneous Sharpe ratio is approximately constant. It is this state-dependence in the indirect utility function that induces hedging demand. The third panel of Figure 1-7 shows hedging demand of the irrational trader. Over a wide range of values of D_t , his hedging demand is non-zero. In particular, it is close to its asymptotic value $\eta(\gamma - 1)/\gamma$ (see Proposition 7), which equals 12.8 for the chosen values of parameters.

What we conclude from this is that convergence of the stock price to a limiting process does not necessarily imply convergence of the traders' portfolio policies to their policies under the limiting price process. Price paths of small probability under the true probability measure can have a significant impact on the traders' portfolio policies. Thus, an intuitive conjecture that convergence in price gives convergence in portfolio policies does not hold in general. This result has important implications for the analysis of long-run survival as we see in the next section.

1.9 Heuristic Partial Equilibrium Analysis of Survival

Although general equilibrium analysis is always desirable, its tractability is often limited. Several authors such as DSSW have relied on heuristic partial equilibrium analysis to study the survival of irrational traders. In this section, we want to examine

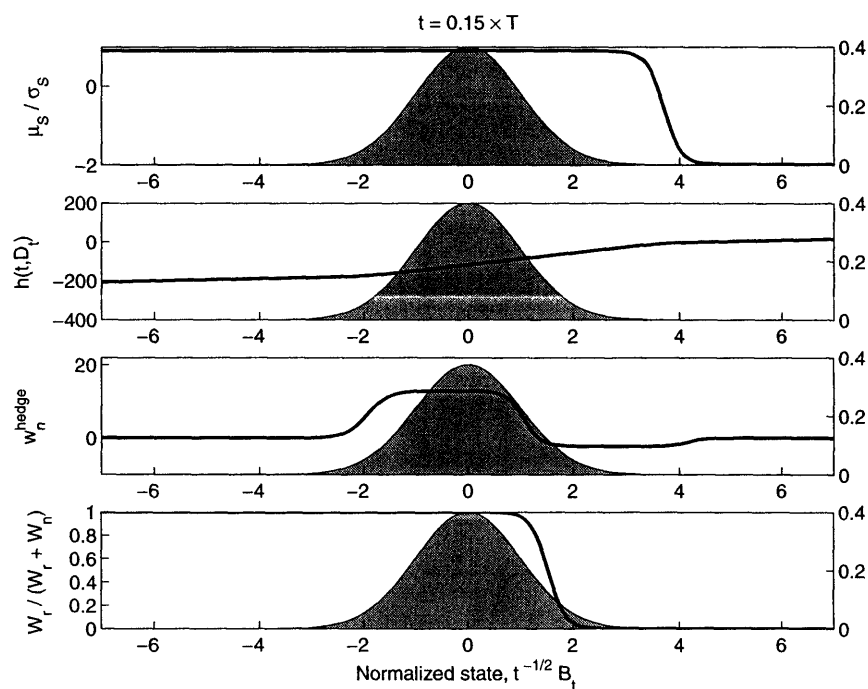


Figure 1-7: The behavior of the economy for the following parameter values: $\mu = 0.12$, $\sigma = 0.18$, $\gamma = 5$, $T = 30$. Also, $\eta = 2\eta^* = 16$, i.e., the irrational trader is strongly optimistic. The time of observation is set at $t = 0.15 \times T$. The horizontal axis in all panels is the normalized state variable, $g_{0,T} = B_T/\sqrt{T}$, which has a standard normal distribution with zero mean and unit variance, shown by the shaded area (vertical axis on the right). The four panels from top to bottom show (i) the instantaneous Sharpe ratio of stock returns, μ_S/σ_S ; (ii) the state dependence of the indirect value function of the rational trader, as captured by the function $h(t, D_t)$ in (1.20); (iii) the portion of the portfolio strategy of the irrational trader attributable to hedging demand, defined as $w_n^{\text{hedge}} = w_n - \mu_S + \eta\sigma_S^2/(\gamma\sigma_S^2)$; (iv) the fraction of the aggregate wealth controlled by the rational agent, $W_r/(W_r + W_n)$.

the limitations of partial equilibrium heuristics in our setting.

The essence of the partial equilibrium argument is to examine a limiting situation when one of the two traders controls most of the aggregate wealth. Following DSSW, the argument then assumes that the infinitesimal trader has no impact on market prices and all traders follow portfolio policies close to those under the limiting prices. If the wealth of the infinitesimally small trader has a higher growth rate under the assumed portfolio policies, his share of wealth will grow over time and he will be able to successfully “invade” the economy. Hence, such traders can survive in the long-run, “in the sense that their wealth share does not drop toward zero in the long run with probability one”.

In our setting, we can easily derive the survival conditions using this partial equilibrium argument. In the limit when the economy is populated only by either the rational trader or the irrational trader, the stock price follows the geometric Brownian motion:

$$dS_t = S_t (\mu_S dt + \sigma_S dB_t). \quad (1.21)$$

If only the rational trader is present, $S_t = S_t^*$ and we have from (1.12) $\mu_S = \gamma\sigma^2$ and $\sigma_S = \sigma$. He invests only in the stock and rate of his wealth growth is given by $\mu_S - \frac{1}{2}\sigma_S^2 = \frac{1}{2}(2\gamma - 1)\sigma^2$.

Suppose now an irrational trader is injected into the economy. Under his belief (given by the measure Q), the drift of the stock price process is $\hat{\mu}_S = \mu_S + \sigma^2\eta$ and the volatility remains at σ . He will chose to invest a fraction $w_n = \hat{\mu}_S / (\gamma\sigma^2) = 1 + \eta/\gamma$ of his wealth in the stock. Thus, the growth rate of the irrational trader’s wealth is $\mu_S - \frac{1}{2}\sigma^2 + \frac{1}{2}\frac{\sigma^2}{\gamma^2}\eta(\gamma\eta^* - \eta)$ where $\eta^* = 2(\gamma - 1)$. The growth rate of wealth of the “invading” irrational trader is higher than that of the dominant rational trader if and only if $0 < \eta < \gamma\eta^*$.

Next, assume that only the irrational trader is dominant. Then, $S_t = S_t^{**}$. Repeating the steps of the previous analysis, the volatility of the limiting stock price remains at σ and the drift becomes $\mu_S = \gamma\sigma^2 - \eta\sigma^2$. The growth rate of the irrational trader’s wealth is $\mu_S - \frac{1}{2}\sigma^2$ while for the rational trader it is $\mu_S - \frac{1}{2}\sigma^2 + \frac{1}{2}\frac{\sigma^2}{\gamma^2}(2\gamma -$

1) $\eta \left(\eta - \frac{\gamma}{2\gamma-1}\eta^* \right)$. The rational trader's portfolio grows faster than the irrational trader's portfolio if and only if $\eta < 0$ or $\eta > \frac{\gamma}{2\gamma-1}\eta^*$.

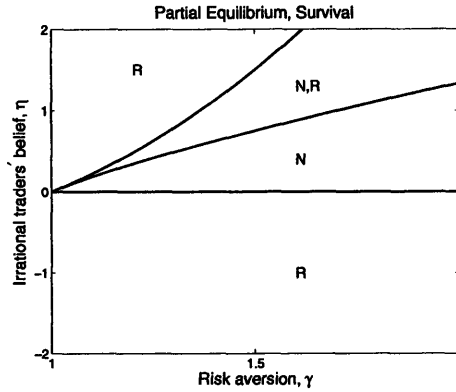


Figure 1-8: The survival of rational and irrational traders for different values of η and γ in partial equilibrium. For each region in the parameter space, we document which of the agents survives in the long run. “R” means that survival of the rational trader is guaranteed inside the region, “N” corresponds to the irrational trader, “N,R” means that both traders survive.

The partial equilibrium analysis thus appears to provide sufficient conditions for long-run survival of both types of traders. In particular, for $\gamma > 1$

$$\begin{aligned}
 0 < \eta < \frac{\gamma}{2\gamma-1}\eta^* &\Rightarrow \text{Irrational trader survives} \\
 \frac{\gamma}{2\gamma-1}\eta^* < \eta < \gamma\eta^* &\Rightarrow \text{Both traders survive} \\
 \eta < 0 \text{ or } \eta > \gamma\eta^* &\Rightarrow \text{Rational trader survives}
 \end{aligned}
 \tag{1.22}$$

For $\gamma = 1$, only the rational trader survives regardless of the value of η . Figure 1-8 summarizes these results. Since $\gamma/(2\gamma - 1) \leq 1$ for $\gamma \geq 1$, η^* belongs to the second region in (1.22).

The survival conditions given in Figure 1-8 clearly differ from the survival conditions from general equilibrium analysis shown in Figure 1-5. The difference occurs when $\frac{\gamma}{2\gamma-1}\eta^* < \eta < \gamma\eta^*$. In particular, partial equilibrium argument predicts survival of both traders for these parameter values while general equilibrium analysis shows the extinction of the irrational trader when $\eta > \eta^*$.

The difference in results from the partial equilibrium argument comes from its

two assumptions: (1) when the irrational trader becomes small in relative wealth, the stock price behaves as if he is absent, and (2) both traders adopt the portfolio policies that would be optimal under that limiting price process. We know from our analysis in Section 1.5 that the first assumption is false in general. But the more direct reason for the discrepancy in survival results is because the second assumption is false. For instance, $\eta^* < \eta < \gamma\eta^*$ corresponds to Case 3 of Proposition 5, in which the stock price is asymptotically the same as in the economy without the irrational trader. In other words, the irrational trader has no significant impact on the current stock price as his wealth becomes negligible. The moments of stock returns converge to the values implied by the partial equilibrium analysis. However, as we have shown in Section 1.8, the irrational trader's portfolio policy differs significantly from what the partial equilibrium analysis assumes. In particular, he does not simply hold the portfolio implied by the limiting price process. This explains the deviations in the conclusions about long-run survival from the heuristic partial equilibrium argument and demonstrates the limitations of partial equilibrium arguments and the importance of equilibrium effects on survival.

1.10 Conclusion

The analysis above has examined the long-run price impact and survival of irrational traders who use persistently wrong beliefs to make their portfolio choices. Using a parsimonious model with no intermediate consumption, we have shown that irrational traders can maintain a persistent influence on prices even after they have lost most of their wealth. Our analysis of conditions for survival of either type of traders further highlights the importance of taking into account the effect that traders have on asset prices.

For tractability, we have confined our analysis to preferences with constant relative risk aversion. Extensions of our analysis to more general preferences are possible and may yield unexpected results. We have also assumed that the rational and

irrational traders differ only in their beliefs but not in their preferences. This allows us to focus on the impact of irrational beliefs on survival and prices. Of course, differences in time and risk preferences can have their own implications for long-run survival. Perhaps more important is the extension of these results to models with intermediate consumption and to alternative preferences. While there is more to be done in this area, it is fair to say that a general message is emerging and is unlikely to be overturned. Namely, survival and price impact are related but distinct concepts and the arguments ignoring such a distinction are unreliable. In our model, irrational traders can survive and even dominate rational traders, but even when they do not survive, they can still have a persistent impact on asset prices.

1.11 Appendix

Proof of Proposition 1

The optimality conditions of the maximization problem in (1.8a) require that

$$C_{r,T} = C_{n,T} (b \xi_T)^{1/\gamma}.$$

Combined with the market clearing condition (1.8b), this implies (1.9a) and (1.9b).

The state price density must be proportional to the traders' marginal utilities. Since we set the interest rate equal to zero, the state price density conditional on the information available at time t is given by

$$\frac{\left(1 + (b \xi_T)^{1/\gamma}\right)^\gamma D_T^{-\gamma}}{\mathbb{E}_t \left[\left(1 + (b \xi_T)^{1/\gamma}\right)^\gamma D_T^{-\gamma} \right]}.$$

The price of any payoff Z_T is therefore given by (1.11).

The individual budget constraint in a dynamically complete market is equivalent to the static constraint that the initial wealth of a trader is equal to the present value of the trader's consumption (e.g., Cox and Huang (1989)). Since the two traders in

our model have identical endowments at time $t = 0$, their budget constraints imply

$$W_{r,0} = \frac{\mathbb{E}_0 \left[D_T^{1-\gamma} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^{\gamma-1} \right]}{\mathbb{E}_0 \left[D_T^{-\gamma} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^{\gamma} \right]} = \frac{\mathbb{E}_0 \left[D_T^{1-\gamma} (b \xi_T)^{\frac{1}{\gamma}} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^{\gamma-1} \right]}{\mathbb{E}_0 \left[D_T^{-\gamma} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^{\gamma} \right]} = W_{n,0}. \quad (1.23)$$

We now verify that $b = e^{\eta\sigma^2(\gamma-1)T}$ satisfies (1.23). Note that

$$D_T^{1-\gamma} = e^{[(1-\gamma)(\mu - \frac{\sigma^2}{2}) + \frac{1}{2}(1-\gamma)^2\sigma^2]T} e^{-\frac{1}{2}(1-\gamma)^2\sigma^2T + (1-\gamma)\sigma B_T}$$

Define a new measure Q , such that $\left(\frac{dQ}{dP}\right)_T = e^{-\frac{1}{2}(1-\gamma)^2\sigma^2T + (1-\gamma)\sigma B_T}$, where P is the original probability measure. Using the translation invariance property of the Gaussian distribution, the random variable $B_T^Q = B_T - (1-\gamma)\sigma T$ is a standard normal random variable under Q . Thus, the equality

$$\mathbb{E}_0 \left[D_T^{1-\gamma} (b \xi_T)^{\frac{1}{\gamma}} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^{\gamma-1} \right] = \mathbb{E}_0 \left[D_T^{1-\gamma} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^{\gamma-1} \right]$$

is equivalent to

$$\mathbb{E}_0^Q \left[\left((\xi_T^Q)^{\frac{1}{\gamma}} \left(1 + (\xi_T^Q)^{\frac{1}{\gamma}} \right)^{\gamma-1} \right) \right] = \mathbb{E}_0^Q \left[\left(1 + (\xi_T^Q)^{\frac{1}{\gamma}} \right)^{\gamma-1} \right]$$

where $\xi_T^Q = \exp(-\frac{1}{2}\sigma^2\eta^2T + \sigma\eta B_T^Q)$. Since the variable B_T^Q is equivalent in distribution to B_T , we can restate the last equality equivalently as

$$\mathbb{E}_0 \left[\xi_T^{\frac{1}{\gamma}} \left(1 + \xi_T^{\frac{1}{\gamma}} \right)^{\gamma-1} \right] = \mathbb{E}_0 \left[\left(1 + \xi_T^{\frac{1}{\gamma}} \right)^{\gamma-1} \right].$$

To verify that the above equality holds, consider a function $F(z)$ defined as

$$F(z) = \mathbb{E}_0 \left[\left(e^{\frac{1}{2\gamma}zT} + e^{-\frac{1}{2\gamma}zT} \xi_T^{\frac{1}{\gamma}} \right)^{\gamma} \right].$$

Changing the order of differentiation and expectation operators, (see Billingsley 1995,

Th. 16.8),

$$F'(z)|_{z=0} = \mathbb{E} \left[\frac{1}{2} \left(1 - \xi_T^{\frac{1}{\gamma}} \right) \left(1 + \xi_T^{\frac{1}{\gamma}} \right)^{\gamma-1} \right].$$

Thus it suffices to prove that $F'(z)|_{z=0} = 0$. Since

$$\mathbb{E}_0 \left[\left(e^{\frac{1}{2\gamma}zT} + e^{-\frac{1}{2\gamma}zT} \xi_T^{\frac{1}{\gamma}} \right)^\gamma \right] = \mathbb{E}_0 \left[\left(e^{\frac{1}{2\gamma}(zT - \frac{1}{2}\eta^2\sigma^2T + \eta\sigma B_T)} + e^{-\frac{1}{2\gamma}(zT - \frac{1}{2}\eta^2\sigma^2T + \eta\sigma B_T)} \right)^\gamma \xi_T^{\frac{1}{2}} \right]$$

if we define a new measure Q so that $\left(\frac{dQ}{dP}\right)_T = e^{-\frac{1}{8}\eta^2\sigma^2T + \frac{1}{2}\eta\sigma B_T}$ and use a change of measure similar to its earlier application in this proof, we find that

$$\mathbb{E}_0 \left[\left(e^{\frac{1}{2\gamma}zT} + e^{-\frac{1}{2\gamma}zT} \xi_T^{\frac{1}{\gamma}} \right)^\gamma \right] = \mathbb{E}_0 \left[\left(e^{\frac{1}{2\gamma}(zT + \eta\sigma B_T)} + e^{-\frac{1}{2\gamma}(zT + \eta\sigma B_T)} \right)^\gamma \right] e^{-\frac{1}{8}\eta^2\sigma^2T}.$$

The symmetry of the distribution of the normal random variable B_T implies that $F(z) = F(-z)$, therefore $F'(z)|_{z=0} = 0$. This verifies that $b = e^{\eta\sigma^2(\gamma-1)T}$.

We now prove that the conditional volatility of stock returns is bounded between σ and $\sigma(1 + |\eta|)$. Define

$$A = e^{\left(\frac{-\eta\sigma^2}{\gamma}\right)(T-t)}, \quad g = e^{-\frac{1}{2}\eta^2\sigma^2\frac{1}{\gamma}T + \sigma^2\eta(\gamma-1)\frac{1}{\gamma}t + \frac{\eta\sigma}{\gamma}B_T}.$$

The stock price can be expressed as

$$S_t = \frac{\mathbb{E}_t \left[D_T^{1-\gamma} \left(1 + (b\xi_T)^{1/\gamma} \right)^\gamma \right]}{\mathbb{E}_t \left[D_T^{-\gamma} \left(1 + (b\xi_T)^{1/\gamma} \right)^\gamma \right]} = e^{(\mu - \sigma^2\gamma)T + (-\frac{1}{2}\sigma^2(1-2\gamma))t} e^{\sigma B_t} \frac{\mathbb{E}_t [(1+g)^\gamma]}{\mathbb{E}_t [(1+gA)^\gamma]}.$$

By Ito's lemma, its volatility σ_{S_t} is given by

$$\sigma_{S_t} = \frac{\partial \ln S_t}{\partial B_t} = \sigma + \eta\sigma \left(\frac{\mathbb{E}_t [(1+gA)^{\gamma-1}]}{\mathbb{E}_t [(1+gA)^\gamma]} - \frac{\mathbb{E}_t [(1+g)^{\gamma-1}]}{\mathbb{E}_t [(1+g)^\gamma]} \right). \quad (1.24)$$

To establish the bounds on stock return volatility, we prove that

$$\frac{\mathbb{E}_t [(1+gA)^{\gamma-1}]}{\mathbb{E}_t [(1+gA)^\gamma]} - \frac{\mathbb{E}_t [(1+g)^{\gamma-1}]}{\mathbb{E}_t [(1+g)^\gamma]} \geq 0 \quad (1.25)$$

for $A \leq 1$ with the opposite inequality for $A \geq 1$. Note that for any twice-differentiable function $F(A, \gamma)$,

$$\frac{\partial}{\partial \gamma} \frac{\partial}{\partial A} \ln(F(A, \gamma)) \geq 0 \Rightarrow \frac{\partial}{\partial A} \ln(F(A, \gamma - 1)) - \frac{\partial}{\partial A} \ln[F(A, \gamma)] \leq 0 \Rightarrow \frac{\partial}{\partial A} \frac{F(A, \gamma - 1)}{F(A, \gamma)} \leq 0.$$

Thus, to prove (1.25), it suffices to show that $\partial^2 \ln(\mathbb{E}_t[(1 + gA)^\gamma]) / \partial A \partial \gamma \geq 0$. The function $(1 + gA)^\gamma$ is log-supermodular in A , g , and γ , since it is positive and its cross-partial derivatives in all arguments are positive. Thus, according to the additivity property of log-supermodular functions (e.g., Athey (2002)), $\mathbb{E}_t[(1 + gA)^\gamma]$ is log-supermodular in A and γ , i.e., $\partial^2 \ln(\mathbb{E}_t[(1 + gA)^\gamma]) / \partial A \partial \gamma \geq 0$.

Because $A > 1$ if and only if $\eta < 0$, we have shown that

$$\eta \left(\frac{\mathbb{E}_t[(1 + gA)^{\gamma-1}]}{\mathbb{E}_t[(1 + gA)^\gamma]} - \frac{\mathbb{E}_t[(1 + g)^{\gamma-1}]}{\mathbb{E}_t[(1 + g)^\gamma]} \right) \geq 0$$

and hence $\sigma_{St} \geq \sigma$.

Because $\left(\frac{\mathbb{E}_t[(1 + gA)^{\gamma-1}]}{\mathbb{E}_t[(1 + gA)^\gamma]} - \frac{\mathbb{E}_t[(1 + g)^{\gamma-1}]}{\mathbb{E}_t[(1 + g)^\gamma]} \right)$ is bounded between -1 and 0 for $\eta < 0$, and between 0 and 1 for $\eta > 0$, we obtain the upper bound from (1.24): $\sigma_{St} \leq \sigma(1 + |\eta|)$.

■

Proof of Proposition 3

We will make use of the following result:

Lemma 1 *Let $N(x)$ denote the cumulative density function of the standard normal distribution: $N(x) \equiv \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{z^2}{2}} dz$. For $x > 0$, $N(x) \leq \frac{1}{2} e^{-\frac{x^2}{2}}$.*

Proof. $N(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{z^2}{2}} dz \leq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{x^2}{2} - \frac{(z-x)^2}{2}} dz = \frac{1}{2} e^{-\frac{x^2}{2}}$. Note that for convenience, we have defined the cumulative density function as the probability above a given value rather than below. ■

Let $t = T/(1 + |\eta|)$ and define $M = \frac{\sqrt{2}-1}{2} \sigma |\eta| t$. According to Lemma 1,

$$\text{Prob} \left[|B_t| \geq M\sqrt{t} \right] = 2N(M\sqrt{t}) \leq e^{-\frac{M^2}{2}t} = e^{-\frac{(\sqrt{2}-1)^2}{2} \sigma^2 \eta^2 t} \leq e^{-\frac{1}{12} \sigma^2 \eta^2 t}.$$

On set $\{|B_t| \leq M\sqrt{t}\}$,

$$\alpha_{n,t} = \frac{\xi_t}{1 + \xi_t} \leq \xi_t \leq e^{-\frac{\sigma^2 \eta^2}{2} t + \sigma |\eta| M t} = e^{-\frac{3-2\sqrt{2}}{2} \sigma^2 \eta^2 t} \leq e^{-\frac{1}{12} \sigma^2 \eta^2 t} \leq \varepsilon.$$

Therefore,

$$\text{Prob}[\alpha_{n,t} \geq \varepsilon] \leq \text{Prob}\left[|B_t| \geq M\sqrt{t}\right] \leq e^{-\frac{1}{12} \sigma^2 \eta^2 t} \leq \varepsilon$$

which establishes the first result of the proposition. The second result follows from the fact that, on the set $\{|B_t| \leq M\sqrt{t}\}$,

$$\frac{S_t}{S_t^*} \leq e^{-\sigma^2 |\eta| (T-t)} \frac{1}{\alpha_{n,t}} \leq e^{-\sigma^2 |\eta| (T-t) + \frac{\sigma^2 \eta^2}{2} t + \sigma |\eta| M \sqrt{t}} \leq e^{-\frac{\sigma^2 \eta^2}{2} t + \sigma |\eta| M t} e^{-\sigma^2 |\eta| T + \sigma^2 |\eta| (1+|\eta|) t}.$$

Given that on the set $\{|B_t| \leq M\sqrt{t}\}$,

$$e^{-\frac{\sigma^2 \eta^2}{2} t + \sigma |\eta| M t} \leq e^{-\frac{1}{12} \sigma^2 \eta^2 t}$$

and since $t = T/(1 + |\eta|)$, $e^{-\sigma^2 |\eta| T + \sigma^2 |\eta| (1+|\eta|) t} \leq 1$ and we conclude that on the set $\{|B_t| \leq M\sqrt{t}\}$, $\frac{S_t}{S_t^*} \leq e^{-\frac{1}{12} \sigma^2 \eta^2 t}$ and hence

$$\text{Prob}\left[1 - \frac{S_t}{S_t^*} \leq 1 - \varepsilon\right] \leq e^{-\frac{1}{12} \sigma^2 \eta^2 t} \leq \varepsilon$$

which concludes the proof of the proposition. ■

Proof of Proposition 4

According to (1.9a) and (1.9b),

$$\frac{C_{n,T}}{C_{r,T}} = (b \xi_T)^{1/\gamma} = \exp\left[\frac{1}{\gamma} \left(-\frac{1}{2} \sigma^2 \eta^2 + \eta \sigma^2 (\gamma - 1)\right) T + \frac{1}{\gamma} \eta \sigma B_T\right].$$

Using the strong Law of Large Numbers for Brownian motion (see Karatzas and Shreve (1991, Sec. 2.9.A)), for any value of σ ,

$$\lim_{T \rightarrow \infty} e^{aT + \sigma B_T} = \begin{cases} 0, & a < 0 \\ \infty, & a > 0 \end{cases}$$

where convergence takes place almost surely. The proposition then follows. ■

Proof of Proposition 5

Our analysis will make use of the following technical result.

Lemma 2 *Consider a stochastic process $X_t = e^{ct + vB_t}$ and a constant $a \geq 0$. Assume that $ac + \frac{1}{2}v^2a^2(1 - \lambda) \neq 0$, $0 \leq \lambda < 1$. Then the limit $\lim_{T \rightarrow \infty} E_t[X_T^a]$ is equal to either zero or infinity almost surely, where we set $t = \lambda T$. The following convergence results hold:*

(i) *(Point-wise convergence)*

$$\lim_{T \rightarrow \infty} \frac{E_t[(1 + X_T)^a]}{1 + E_t[X_T^a]} = 1. \quad (1.26)$$

(ii) *(Convergence of moments)*

$$\lim_{T \rightarrow \infty} \frac{\text{mean}_t E_t[(1 + X_T)^a]}{\text{mean}_t (1 + E_t[X_T^a])} = 1, \quad \lim_{T \rightarrow \infty} \frac{\text{vol}_t E_t[(1 + X_T)^a]}{\text{vol}_t (1 + E_t[X_T^a])} = 1 \quad (1.27)$$

where $\text{mean}_t f_t$ and $\text{vol}_t f_t$ denote the instantaneous mean and standard deviation of the process $\ln f_t$ respectively.

Proof of Lemma 2

(i) Consider the conditional expectation

$$E_t[X_T^a] = \exp \left[acT + \frac{1}{2}v^2a^2(1 - \lambda)T + avB_t \right]. \quad (1.28)$$

The limit of $E_t[X_T^a]$ is equal to zero if $ac + \frac{1}{2}v^2a^2(1 - \lambda) < 0$ and equal to infinity if the opposite inequality holds (according to the strong Law of Large Numbers for Brownian motion, see Karatzas and Shreve, 1991, Sec. 2.9.A).

Because the function $acT + \frac{1}{2}v^2a^2(1 - \lambda)T$ is convex in a and equal to zero when $a = 0$, we find that for $a \geq 1$

$$E_t[X_T^a] \rightarrow \infty \Rightarrow \frac{E_t[X_T^z]}{E_t[X_T^a]} \rightarrow 0, \quad \forall z \in (0, a) \quad (1.29a)$$

$$E_t[X_T^a] \rightarrow 0 \Rightarrow E_t[X_T^z] \rightarrow 0, \quad \forall z \in (0, a). \quad (1.29b)$$

We prove the result of the lemma separately for six regions covering the entire parameter space.

Case 1: $0 \leq a \leq 1, E_t[X_T^a] \rightarrow \infty$.

If $X_T \leq 1$, $(X_T + 1)^a \leq 2^a$, while if $X_T \geq 1 \Rightarrow (X_T + 1)^a - X_T^a \leq aX_T^{a-1} \leq a$ since $(X_T + 1)^a$ is concave and $a - 1 \leq 0$. Therefore, $X_T^a \leq (1 + X_T)^a \leq X_T^a + 2^a + a$, and hence $\lim_{T \rightarrow \infty} E_t[(1 + X_T)^a]/E_t[X_T^a] = 1$, which implies $\lim_{T \rightarrow \infty} E_t[(1 + X_T)^a]/(1 + E_t[X_T^a]) = 1$.

Case 2: $1 \leq a \leq 2, E_t[X_T^a] \rightarrow \infty$.

By the mean value theorem, $(1 + X_T)^a = X_T^a + a(w + X_T)^{a-1}$ for some $w \in [0, 1]$. Using the analysis of case 1, $(w + X_T)^{a-1} \leq (1 + X_T)^{a-1} \leq X_T^{a-1} + 2^{a-1} + a - 1$, which, combined with (1.29a), implies that $\lim_{T \rightarrow \infty} E_t[(1 + X_T)^a]/E_t[X_T^a] = 1$ and the main result follows.

Case 3: $2 \leq a, E_t[X_T^a] \rightarrow \infty$.

By the mean value theorem, $(1 + X_T)^a = X_T^a + a(w + X_T)^{a-1}$ for some $w \in [0, 1]$. By Jensen's inequality, $[(1 + X_T)/2]^{a-1} \leq (1 + X_T^{a-1})/2$. Thus,

$$0 \leq (w + X_T)^{a-1} \leq (1 + X_T)^{a-1} \leq 2^{a-2} + 2^{a-2}X_T^{a-1}$$

which, combined with (1.29a) implies that $\lim_{T \rightarrow \infty} E_t[(1 + X_T)^a]/E_t[X_T^a] = 1$ and the main result follows.

Case 4: $0 \leq a \leq 1, E_t[X_T^a] \rightarrow 0$:

If $X_T \leq 1$, $(1 + X_T)^a \leq 1 + X_T \leq 1 + X_T^a$, while if $X_T \geq 1$, $(1 + X_T)^a \leq X_T^a + a \leq 1 + X_T^a$ since $(1 + X_T)^a$ is concave. Thus, $1 \leq (1 + X_T)^a \leq 1 + X_T^a$ and therefore $\lim_{T \rightarrow \infty} E_t[(1 + X_T)^a] = 1$, which implies the main result.

Case 5: $1 \leq a \leq 2, E_t[X_T^a] \rightarrow 0$.

By the mean value theorem, $(1 + X_T)^a = 1 + aX_T(1 + wX_T)^{a-1}$ for some $w \in [0, 1]$. Further, $X_T(1 + wX_T)^{a-1} \leq X_T(1 + X_T)^{a-1} \leq X_T(X_T^{a-1} + 2^{a-1} + a - 1)$, using the same argument as in case 1. Since $\lim_{T \rightarrow \infty} E_t[X_T^a] = 0$, according to (1.29b), $\lim_{T \rightarrow \infty} E_t[X_T] = 0$ and hence $\lim_{T \rightarrow \infty} E_t[(1 + X_T)^a] = 1$.

Case 6: $2 \leq a, E_t[X_T^a] \rightarrow 0$.

By the mean value theorem, $(1 + X_T)^a = 1 + aX_T(1 + wX_T)^{a-1}$ for some $w \in [0, 1]$. Further, $X_T(1 + wX_T)^{a-1} \leq X_T(1 + X_T)^{a-1} \leq 2^{a-2}X_T + 2^{a-2}X_T^a$ by Jensen's inequality. Since $\lim_{T \rightarrow \infty} E_t[X_T^a] = 0$, according to (1.29b), and $\lim_{T \rightarrow \infty} E_t[X_T] = 0$ and hence $\lim_{T \rightarrow \infty} E_t[(1 + X_T)^a] = 1$.

(ii) Since the conditional expectations $E_t[(1 + X_T)^a]$ and $E_t[1 + X_T^a]$ are martingales, they have zero drift for all values of T and t . By Ito's lemma, convergence of the first moments of the natural logarithms of the same processes follows from convergence of the second moments.

We now establish convergence of volatility of the process $E_t[(1 + X_T)^a]$. According to Ito's lemma, one must show that

$$\lim_{T \rightarrow \infty} \frac{\partial \ln E_t[(1 + X_T)^a] / \partial B_t}{\partial \ln(1 + E_t[X_T^a]) / \partial B_t} = 1, \quad \forall a \geq 0.$$

Given (1.28), it suffices to prove that $\lim_{T \rightarrow \infty} \partial \ln E_t[(1 + X_T)^a] / \partial B_t = 0$ if $\lim_{T \rightarrow \infty} E_t[X_T^a] = 0$ and $\lim_{T \rightarrow \infty} \partial \ln E_t[(1 + X_T)^a] / \partial B_t = av$ if $\lim_{T \rightarrow \infty} E_t[X_T^a] = \infty$.

First, changing the order of differentiation and expectation operators (see Billingsley 1995, Th. 16.8),

$$\frac{\partial \ln E_t[(1 + X_T)^a]}{\partial B_t} = av \frac{E_t[X_T(1 + X_T)^{a-1}]}{E_t[(1 + X_T)^a]} = av \left(1 - \frac{E_t[(1 + X_T)^{a-1}]}{E_t[(1 + X_T)^a]} \right).$$

Furthermore, according to part (i),

$$\frac{\mathbb{E}_t[(1 + X_T)^{a-1}]}{\mathbb{E}_t[(1 + X_T)^a]} \sim \frac{\mathbb{E}_t[(1 + X_T)^{a-1}]}{1 + \mathbb{E}_t[X_T^a]}. \quad (1.30)$$

Assume $a \geq 1$. As we have shown in case 1 of the proof of part (i), $X_T^{a-1} \leq (1 + X_T)^{a-1} \leq X_T^{a-1} + 2^{a-1} + a - 1$. If $\mathbb{E}_t[X_T^a] \rightarrow \infty$, according to (1.29a), $\mathbb{E}_t[X_T^{a-1}]/\mathbb{E}_t[X_T^a] \rightarrow 0$, which yields

$$\lim_{T \rightarrow \infty} \frac{\partial \ln \mathbb{E}_t[(1 + X_T)^a]}{\partial B_t} = av.$$

Similarly, if $\mathbb{E}_t[X_T^a] \rightarrow 0$, then, according to (1.29b), $\lim_{T \rightarrow \infty} \mathbb{E}_t[X_T^{a-1}] = 0$, which, according to part (i), implies that $\lim_{T \rightarrow \infty} \mathbb{E}_t[(1 + X_T)^{a-1}] = 1$ and

$$\lim_{T \rightarrow \infty} \frac{\partial \ln \mathbb{E}_t[(1 + X_T)^a]}{\partial B_t} = 0.$$

Next, consider the case of $0 < a < 1$. If $\mathbb{E}_t[X_T^a] \rightarrow \infty$, because $\mathbb{E}_t[(1 + X_T)^{a-1}] \leq 1$, (1.30) implies $\lim_{T \rightarrow \infty} \partial \ln \mathbb{E}_t[(1 + X_T)^a]/\partial B_t = av$.

Suppose that $\lim_{T \rightarrow \infty} \mathbb{E}_t[X_T^a] = 0$. By Markov's inequality, for any $\epsilon > 0$, $P_t[X_T > \epsilon] \leq \mathbb{E}_t[X_T^a]/\epsilon^a \rightarrow 0$. Similarly, $P_t[X_T < \epsilon] \leq \mathbb{E}_t[(1 + X_T)^{a-1}]/(1 + \epsilon)^{a-1}$. Thus, $1 \geq \mathbb{E}_t[(1 + X_T)^{a-1}] \geq P_t[X_T < \epsilon](1 + \epsilon)^{a-1}$, and $\liminf_{T \rightarrow \infty} \mathbb{E}_t[(1 + X_T)^{a-1}] \geq (1 + \epsilon)^{a-1}$ for any $\epsilon > 0$. This implies that $\lim_{T \rightarrow \infty} \mathbb{E}_t[(1 + X_T)^{a-1}] = 1$ and $\lim_{T \rightarrow \infty} \partial \ln \mathbb{E}_t[(1 + X_T)^a]/\partial B_t = 0$. ■

We establish the long-run behavior of S_t for the case when $\gamma > 1$ and $0 < \eta < \eta^* = 2(\gamma - 1)$. The results for all other regions in the parameter space can be obtained similarly.

The equilibrium stock price and the ratio of the individual wealth processes are given by

$$S_t = \frac{\mathbb{E}_t \left[D_T^{1-\gamma} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^\gamma \right]}{\mathbb{E}_t \left[D_T^{-\gamma} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^\gamma \right]}, \quad \frac{W_{r,t}}{W_{n,t}} = \frac{\mathbb{E}_t \left[D_T^{1-\gamma} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^{\gamma-1} \right]}{\mathbb{E}_t \left[D_T^{1-\gamma} (b \xi_T)^{\frac{1}{\gamma}} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^{\gamma-1} \right]}.$$

We therefore need to characterize the long-run behavior of the following two quantities:

$$E^{(1)} \equiv \mathbb{E}_t \left[D_T^{1-\gamma} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^\gamma \right], \quad E^{(2)} \equiv \mathbb{E}_t \left[D_T^{-\gamma} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^\gamma \right].$$

Consider the first expression,

$$E^{(1)} = \mathbb{E}_t \left[D_T^{1-\gamma} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^\gamma \right] = D_t^{1-\gamma} \mathbb{E}_t \left[\left(\frac{D_T}{D_t} \right)^{1-\gamma} \left(1 + \left(b \xi_t \frac{\xi_T}{\xi_t} \right)^{\frac{1}{\gamma}} \right)^\gamma \right].$$

Given the aggregate dividend process,

$$\left(\frac{D_T}{D_t} \right)^{1-\gamma} = e^{(T-t)(\mu(1-\gamma) - \frac{1}{2}\sigma^2(1-\gamma)\gamma)} e^{-\frac{1}{2}(1-\gamma)^2\sigma^2(T-t) + (1-\gamma)\sigma(B_T - B_t)}.$$

As in the proof of Proposition 1, we introduce a new measure Q with the Radon-Nikodym derivative $\left(\frac{dQ}{dP} \right)_t = e^{-\frac{1}{2}(1-\gamma)^2\sigma^2(T-t) + (1-\gamma)\sigma(B_T - B_t)}$. By Girsanov's theorem, $B_T - B_t = B_T^Q - B_t^Q - (1-\gamma)\sigma(T-t)$, where B_t^Q is a Brownian motion under the measure Q . Using the expression for b from Proposition 1, $b = e^{T(\gamma-1)\sigma^2\eta}$, we find

$$E^{(1)} = e^{T(\mu(1-\gamma) - \frac{1}{2}\sigma^2(1-\gamma)\gamma) + t(-\frac{1}{2}\sigma^2(1-\gamma)^2) + B_t(\sigma(1-\gamma))} \mathbb{E}_t^Q \left[\left(1 + e^{(-\frac{1}{2}\eta^2\sigma^2\frac{1}{\gamma})T + (\frac{1}{\gamma}(\gamma-1)\sigma^2\eta)t + \frac{\eta\sigma}{\gamma}B_T^Q} \right)^\gamma \right].$$

We will omit the superscript Q , since the distribution of B_t^Q under the measure Q is the same as the distribution of B_t under the original measure P .

Using the assumption that $t = \lambda T$, define

$$X_T = e^{(-\frac{1}{2}\eta^2\sigma^2\frac{1}{\gamma} + (1-\lambda)\frac{1}{\gamma}(\gamma-1)\sigma^2\eta)T + \frac{\eta\sigma}{\gamma}B_T}.$$

We now apply the result of lemma 2, with

$$c = -\frac{1}{2}\eta^2\sigma^2\frac{1}{\gamma} + (1-\lambda)\frac{1}{\gamma}(\gamma-1)\sigma^2\eta, \quad v = \frac{\eta\sigma}{\gamma}, \quad a = \gamma.$$

Since we are assuming $\gamma > 1$ and $0 < \eta < 2(\gamma-1)$, $\lim_{T \rightarrow \infty} \mathbb{E}_t[X_T^a] = \infty$. According

to lemma 2,

$$\mathbb{E}_t \left[D_T^{1-\gamma} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^\gamma \right] \sim e^{(\mu(1-\gamma) - \frac{1}{2}\sigma^2(1-\gamma)\gamma)T + (-\frac{1}{2}\sigma^2(\eta+1-\gamma)^2)t + \sigma(\eta+1-\gamma)B_t}.$$

We next examine $E^{(2)}$. Using a similar change of measure, we find

$$E^{(2)} = e^{(-\mu\gamma + \frac{1}{2}\sigma^2(1+\gamma)\gamma)T + (-\frac{1}{2}\sigma^2\gamma^2)t + (-\sigma\gamma)B_t} \mathbb{E}_t \left[\left(1 + e^{(-\sigma^2\eta\frac{1}{\gamma} - \frac{1}{2}\eta^2\sigma^2\frac{1}{\gamma})T + \sigma^2\eta t + \frac{\eta\sigma}{\gamma}B_T} \right)^\gamma \right].$$

We apply lemma 2, setting $X_T = e^{cT + vB_T}$ and

$$c = -\sigma^2\eta\frac{1}{\gamma} - \frac{1}{2}\eta^2\sigma^2\frac{1}{\gamma} + (1-\lambda)\sigma^2\eta, \quad v = \frac{\eta\sigma}{\gamma}, \quad a = \gamma.$$

The value of $\lim_{T \rightarrow \infty} \mathbb{E}_t[X_T^a]$ depends on the exact combination of the model parameters. In particular,

$$\lim_{T \rightarrow \infty} \mathbb{E}_t[X_T^a] = \begin{cases} \infty, & -2\eta + \lambda(2\gamma\eta - \eta^2) > 0, \\ 0, & -2\eta + \lambda(2\gamma\eta - \eta^2) < 0, \end{cases}$$

(see the proof of lemma 2, part (i)). Define $\lambda_S \equiv \frac{2}{2\gamma - \eta}$. Note that, because $\gamma > 1$ and $0 < \eta < 2(\gamma - 1)$, $0 < \lambda_S < 1$. Then, $\lim_{T \rightarrow \infty} \mathbb{E}_t[X_T^a] = \infty$ if $\lambda > \lambda_S$ and $2\gamma\eta - \eta^2 > 0$ or if $\lambda < \lambda_S$ and $2\gamma\eta - \eta^2 < 0$, and the limit is equal to zero otherwise. By lemma 2, if $\lim_{T \rightarrow \infty} \mathbb{E}_t[X_T^a] = \infty$,

$$\mathbb{E}_t \left[D_T^{-\gamma} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^\gamma \right] \sim e^{T(-\mu\gamma + \frac{1}{2}\sigma^2(1+\gamma)\gamma - \sigma^2\eta) + (-\frac{1}{2}\sigma^2(\eta-\gamma)^2)t + \sigma(\eta-\gamma)B_t}$$

while if $\lim_{T \rightarrow \infty} \mathbb{E}_t[X_T^a] = 0$, then

$$\mathbb{E}_t \left[D_T^{-\gamma} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^\gamma \right] \sim e^{(-\mu\gamma + \frac{1}{2}\sigma^2(1+\gamma)\gamma)T + (-\frac{1}{2}\sigma^2\gamma^2)t + (-\sigma\gamma)B_t}.$$

Using our definition of λ_S , we re-state these results as

$$\mathbb{E}_t \left[D_T^{-\gamma} \left(1 + (b \xi_T)^{\frac{1}{\gamma}} \right)^\gamma \right] \sim \begin{cases} e^{(-\mu\gamma + \frac{1}{2}\sigma^2(1+\gamma)\gamma)T + (-\frac{1}{2}\sigma^2\gamma^2)t - \sigma\gamma B_t}, & 0 \leq \lambda < \lambda_S \\ e^{(-\mu\gamma + \frac{1}{2}\sigma^2(1+\gamma)\gamma - \sigma^2\eta)T + (-\frac{1}{2}\sigma^2(\eta-\gamma)^2)t + \sigma(\eta-\gamma)B_t}, & \lambda_S < \lambda \leq 1 \end{cases}$$

Having established the behavior of both the numerator and the denominator of the expression for the stock price, we have proven the limiting result for the stock price itself. According to part (ii) of lemma 2, not only the stock price, but also the mean and volatility of returns behave according to the asymptotic expressions of Proposition 5 in the limit of the economy horizon T approaching infinity. ■

Proof of Proposition 7

When the financial markets are dynamically complete and there is a single source of uncertainty (driven by a Brownian motion), the fraction of the agent's wealth invested in stock can be computed as a ratio of the instantaneous volatility of the agent's wealth to the instantaneous volatility of the cumulative stock return process. Proposition 5 and (Kogan, Ross, Wang, and Westerfield, 2003, Proposition 8) provide expression for the long-run behavior of the volatility of stock returns and individual wealth processes, from which the expression for portfolio holdings follow immediately. To decompose the portfolio holdings of the rational trader into as a sum of the myopic and hedging demands, we compute the hedging demand as $\mu_S/(\gamma\sigma_S^2)$, where μ_S and σ_S are the drift and the diffusion coefficients of the stock return process. The difference between the total portfolio holdings and the myopic component define the agent's hedging demand. For the irrational trader, the calculations are analogous, except the myopic demand is given by $\hat{\mu}_S/(\gamma\sigma_S^2) = (\mu_S + \eta\sigma\sigma_S)/(\gamma\sigma_S^2)$, where $\hat{\mu}_S$ is the expected stock return as perceived by the irrational trader. ■

Bibliography

- [1] Alpert, M; Raiffa, H. 1959, A progress report on the training of probability assessors. Reprinted in D. Kahneman, P. Slovic, and A. Tversky (eds.) *Judgement under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press, 1982.
- [2] Athey, Susan. 2002, Monotone Comparative Statics Under Uncertainty, *Quarterly Journal of Economics*, 117 (1), 187–223.
- [3] Benartzi, Shlomo; Thaler, Richard H. 2001, Naive Diversification Strategies in Defined Contribution Saving Plans, *American Economic Review*, 91(1), 79–98.
- [4] Billingsley, Patrick. 1995, Probability and Measure, *Wiley-Interscience*, Third Edition.
- [5] Black, Fischer. 1986, Noise, *Journal of Finance*, 41 (3), 529–43.
- [6] Black, Fischer; Scholes, Myron S. 1973, The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 81 (3), 637–54.
- [7] Blume, Lawrence; Easley, David. 2001, If You're So Smart, Why Aren't You Rich? Belief Selection in Complete and Incomplete Markets, Cowles Foundation Discussion Papers, 1319.
- [8] Cox, John; Huang, Chi-fu. 1989, Optimal Consumption and Portfolio Policies when Asset Prices Follow a Diffusion Process, *Journal of Economic Theory*, 49, 33–83.

- [9] De Long, J. Bradford; Shleifer, Andrei; Summers, Lawrence H.; Waldman, Robert J. 1990, Noise Trader Risk in Financial Markets, *The Journal of Political Economy*, 98 (4), 703–38.
- [10] De Long, J. Bradford; Shleifer, Andrei; Summers, Lawrence H.; Waldman, Robert J. 1991, The Survival of Noise Traders in Financial Markets, *Journal of Business*, 64 (1), 1–19.
- [11] Duffie, Darrell; Huang, Chi-fu. 1986, Implementing Arrow-Debreu Equilibria by Continuous Trading of Few Long-Lived Securities, *Econometrica*, 53 (6), 1337–56.
- [12] Figlewski, Stephen. 1978, Market Efficiency in a Market with Heterogeneous Information, *Journal of Political Economy*
- [13] Friedman, Milton. 1953, The Case for Flexible Exchange Rates, *Essays in Positive Economics*, University of Chicago Press.
- [14] Hakansson, Nils H. 1971, Capital Growth and the Mean-Variance Approach to Portfolio Selection, *Journal of Financial and Quantitative Analysis*, 6 (1), 517–557.
- [15] Harrison, J. Michael; Kreps, David M. 1979, Martingales and Arbitrage in Multiperiod Securities Markets, *Journal of Economic Theory*, 2 (3), 381–408.
- [16] Kahneman, Daniel; Tversky, Amos. 1979, Prospect Theory: An Analysis of Decision under Risk, *Econometrica*, 47(2), 263–91.
- [17] Karatzas, Ioannis; Shreve, Steven E. 1991, Brownian Motion and Stochastic Calculus, *Springer-Verlag New York, Inc.*, Second Edition.
- [18] Kogan, Leonid; Ross, Steven; Wang, Jiang; Westerfield, Mark 2003, The Survival and Price Impact of Irrational Traders, NBER working paper 9434.

- [19] Lucas, Robert E. 1978, Asset Prices in an Exchange Economy, *Econometrica*, 46 (6), 1429–45.
- [20] Odean, Terrance. 1998, Are Investors Reluctant to Realize Their Losses?, *Journal of Finance*, 53(5), 1775–98.
- [21] Radner, Roy. 1972, Existence of Equilibrium of Plans, Prices, and Price Expectations in a Sequence of Markets, *Econometrica*, 40 (2), 289–303.
- [22] Sandroni, Alvaro. 2000, Do Markets Favor Agents Able to Make Accurate Predictions?, *Econometrica*, 68 (6), 1303–41.
- [23] Wang, Jiang. 1996, The Term Structure of Interest Rates in a Pure Exchange Economy with Heterogeneous Investors, *Journal of Financial Economics*, 41 (1), 75–110.

Chapter 2

Market Composition and Equity Market Formation

2.1 Introduction

Although investment banks have long maintained that the initial ownership structure of a firm is critical to its valuation, academics have paid significantly less attention to this factor in their analysis of the pricing and performance of Initial Public Offerings. In this paper, I argue that security allocation by investment banks materially influences the proceeds that firms receive from IPOs because a firm's value depends on which investors own its shares. A large noise trader presence in the aftermarket will undermine the willingness of any investor to pay a high price for the firm's shares. Conversely, when investment banks can effectively induce rational trader participation so that a high-quality aftermarket is created, investors will pay a correspondingly higher price for shares offered in the IPO. Because price discretion – underpricing – is the tool by which firms can limit relative noise trader participation, those firms are willing to leave money on the table when they go public.¹

¹The average first day return on IPOs from 1980 through 2001 was 18.8%. In 1999 and 2000, the two years at the height of the recent stock market bubble, the total money left on the table due to underpricing exceeded \$65 billion. For a summary of the IPO literature, please see Jenkinson and Ljungqvist (2001) and Ritter and Welch (2002).

I model the aftermarket in a multi-period setting with two agents, one rational and one noise, who both have short time horizons. Agents must pay an entry cost in order to participate in the stock market. The result is entry equilibria with varying market composition and levels of total participation. After entry, my model is very similar to that of De Long et. al. (1990) (DSSW). Noise traders have incorrect and time-varying beliefs about the dividend which they attempt to exploit once they enter, while rational traders will attempt to exploit the mispricing that noise traders create.² The main difference between my aftermarket model and that of DSSW is that, as a result of the participation decision, noise traders have an unambiguously negative effect on expected price. Moreover, this is true even when the noise traders are on average optimistic. The apparently counter-intuitive result comes from a very strong participation effect: noise traders act to reduce overall participation and market depth and so any risk is born by a smaller number of investors.³ The pricing effect does not rely on changing the level of noise trader risk in the future, it is purely a result of current period participation. Agents, when considering their willingness to invest, will premise their valuation on the characteristics of the trading market that they expect to be created. As a result, the participation decision creates an externality: by choosing to enter, noise traders make entry less desirable for other agents.

With endogenous participation, the noise trader presence in the aftermarket will be persistent. This is because noise trader risk exerts a different effect on the entry decisions of noise and rational traders. Noise traders, perceiving the stock to be under- or over- valued according to their level of optimism, will always perceive a smaller welfare loss to increased risk than will rational traders, who more accurately perceive the stock's value. Again, there is a participation externality: noise traders

²I will show in section 2.2.1 that noise trader beliefs are equivalent to a type of hedging demand. Throughout most of this article, I will use the beliefs interpretation, but the relevant fact is that noise traders have a noisy private value for the stock.

³Because investors believe they hold special information about the value of the stock – noise traders perceive the stock to be over- or under- valued while rational traders realize that the noise traders are incorrect – they will be willing to hold undiversified portfolios. Therefore investors will be sensitive to idiosyncratic risk, and a smaller investor base means the expected returns of the stock must be higher and the price correspondingly lower.

in one period create the very risk that induces their entry in another. Because a stock's price will reflect the discounted value of all future noise trader risk, noise trader persistence amplifies their negative effect on prices. The presence of many noise traders will have persistent effects in the aftermarket and will lower the firm's expected proceeds from an IPO. Firms will take this into account when determining their ownership structure.

Book-building⁴ and other methods that market IPO securities can use discretionary pricing to limit relative noise trader participation exactly because of the different valuation that noise traders have for the stock. Because the noise traders are optimistic and can make their bidding contingent on their private signal, they have an option on their private value. Noise traders gain from both the potential exercise of this option and from any underpricing that may exist. Rational traders, on the other hand, receive value only from underpricing but they are more likely to bid once they have decided to participate because they have no private value and so cannot receive a low signal. By controlling the issue price, the investment bank can determine the relative value of the agents' payoffs and so determine their willingness to participate. To understand the mechanism, consider two polar cases: if the allocation price is above the fair price, only very optimistic noise traders will participate, while if the price is low enough, all traders will wish to bid. Thus, increasing the amount of underpricing can increase the relative participation of rational traders.

In my model, share allocation by investment banks will dominate open auctions as a method of equity market formation, even if an auction would eliminate the "underpricing" of shares in IPOs⁵. The underpricing of shares that normally accompanies

⁴Benveniste and Spindt (1989) is the seminal model of book-building. They describe a situation in which the investment bank uses its discretionary power over prices to induce informed investors to reveal their information. As a result, the issuer is able to reduce the adverse selection problem inherent in having informed and uninformed investors. Benveniste and Wilhelm (1990) and Spatt and Srivastava (1991) further develop models in which the investment bank elicits information through underpricing. My model relies in the ability of the investment bank to set prices, but I do not use asymmetric information, and I do not claim that I have created an optimal mechanism. Instead, I show that firms will attempt to influence the composition of their investors and that they can affect that composition using their discretion over prices.

⁵Derrien and Womack (2003) show that for French IPOs, "the auction mechanism is associated

delegated share allocation and the book building procedure is simply a cost of attracting a desirable set of initial public stockholders to the firm. The result is a higher quality after-market for the firm's shares, a higher open market valuation, and higher *net* IPO proceeds to the issuer. Importantly, this does not rely on a company coming back to the equity markets at a later date – a firm will care about the aftermarket even if the firm will never again issue equity because its prospective buyers care about the aftermarket. These advantageous pricing implications provide an explanation for the popularity of book-building procedures and what would otherwise seem to be an irrational willingness to leave money on the table when taking a firm public.

My general approach is to take the tools of asset pricing to the question of pricing shares in an IPO. While my application is a corporate finance problem, the more fundamental analysis has implications for pricing behavior in a wide range of markets. This is because any security issuer must account for what type of secondary market exists for its securities, and any investor must account for the quality of the market when he or she sells an asset. In fact, my paper has two asset pricing results: market composition can appear to be persistent when traders can choose their level of participation – noise traders have persistent effects – and noise traders, even optimistic noise traders, lower expected asset values. My third result is a corporate finance result: because more noise traders will lower a firm's stock's price, the firm should use an allocation procedure that can effectively limit relative noise trader participation. Moreover, that procedure should not fully adjust to market movements or late increases in investor demand.

At the most general level, my analysis implies that open auctions will not be the best mechanism for marketing securities. And, in fact, within and across countries, book-building methods and fixed price offerings have been gaining market share in IPOs at the expense of open auction formats (Sherman (2002)). However, much of the book-building literature has focused on asymmetric information as the driv-

with less underpricing and lower variance of underpricing”.

ing force behind the mechanisms that dominate IPO allocations.⁶ I do not use any form of asymmetric information in my modelling, and so I am able to offer a new understanding of the security issuing process without excluding the successes that the asymmetric information approach has garnered. I show that investor composition, particularly the effects of noise traders, can explain the prevalence of allocation procedures that give the issuer pricing discretion.

Lastly, because I am able to show that a stock's investor type will influence its price behavior and its expected price, IPOs are not the only possible application. Although I focus on IPOs as a question of great interest, my work could be extended to other firm equity decisions such as Seasoned Equity Offerings, share repurchases, and the like. Any time a firm makes a major change in the structure of the claims on its value, it must take into account the effect that its new investors will have on firm value. So for large share repurchases and SEOs, a firm may want to use price discretion to influence its ownership composition.

My paper will proceed as follows: In section 2.2 I set up and solve a model of the aftermarket, focusing on a two period setting in sections 2.2.1, 2.2.2, and 2.2.3. I then look at a slightly longer duration model in section 2.2.4. In section 2.3, I address the question of how a firm can create a desirable aftermarket for its shares, first looking at the investor's problem in section 2.3.2 and then the firm's problem in section 2.3.3. In section 2.4 I extend my model of the aftermarket to a more dynamic system, and I then show how to create a high quality aftermarket in that setting. I end the aftermarket, IPO, and extension sections with short discussions of the applicability of the intuitions I have described, as well as a discussion of related empirical work.

⁶In addition to the book-building articles mentioned above, Sherman and Titman (2002) derive an equilibrium level of underpricing that will induce investors to acquire information. Additionally, the cascade model of Welch (1992) illustrates how an issuer would wish to control the flow of information during an IPO.

A Small Sample of Related Work

My model of the aftermarket is part of the noise trader literature started by De Long et. al (1990). They focus on the fundamental role of noise and biases in asset markets, whereas I build on that by asking what happens when agents make participation decisions. The differences between my results and theirs, especially my result that noise traders will decrease the expected stock price, all derive from the participation decision. Another model that analyzes participation decisions in the presence of noise traders is that of Jeanne and Rose (2002). However, their setting is exchange rates and they model both noise traders and entry in a different way. The main result is that by using an exchange rate band, a government can reduce the effect of noise trader risk on its currency. Spiegel (1998) finds that in the presence of stochastic supply that stocks can have high and low volatility equilibria. However, noise in his model comes from exogenous shocks rather than from investors who trade.

Ellul and Pagano (2003) and Booth and Chua (1996) have results that resemble my own in that investors are concerned with expected liquidity, and that this liquidity will be inversely related to the “money left on the table” by the issuer. My model differs from theirs because of my focus on market composition rather than liquidity and because I can make predictions about how market movements and increases in demand will affect the issue price.

Two recent papers that share some of my predictions and which make use of behavioral assumptions are those of Loughran and Ritter (2002) and Ljungqvist, Nanda, and Singh (2003). Loughran and Ritter propose a prospect theory explanation of the willingness of firms to leave money on the table in a IPO. Because more money is left on the table when the issuer sees his wealth increase, issuers are more willing to tolerate the loss of revenue than they would be if the money was left on the table when the issuer saw his wealth decline. Thus a bargaining equilibrium allows for underpricing. Furthermore, they can also explain the partial adjustment finding. As with underpricing, their explanation relies on the covariance of changes in wealth with changes in public information. My model is very different in structure, and I

address a different question: why firms do not use open auctions. One might think that if a firm’s biases and behavioral conditions prevented it from obtaining the highest possible proceeds, it would be willing to engage in some type of open auction procedure where these biases might be less actionable. This would be especially true if, as Derrien and Womack (2003) suggest, one of the reasons that auctions show less underpricing than other mechanisms is that they more easily incorporate public news.

Ljungqvist, Nanda, and Singh, on the other hand, argue that underpricing in a “hot” IPO market can be explained in the presence of sentiment investors as a premium that rational agents require to be willing to hold the stock over a period of time that may see the collapse of the “hot” market. To do this, they require that investors face a kind of incentive compatibility constraint – penalties for flipping – or that they are in collusion with the investment bank. But in practice penalty bids for “flipping” are rarely assessed against institutions (Aggarwal (2003)). My model differs from theirs in that I do not limit the selling ability of agents in the aftermarket, and I do not restrict my setting to that of “hot” markets.

2.2 A Model of the Aftermarket

2.2.1 Assets and Agents

The economy has two assets. The first is a storage technology for the consumption good that risklessly returns R for each unit stored. This asset can also be interpreted as a bond in perfectly elastic supply with a riskless gross rate of return R . Thus, $R - 1$ is the riskless rate.

The second asset is a long lived stock in fixed supply S that pays an i.i.d. dividend every period:

$$\tilde{d}_t \sim N(\bar{d}, \sigma_d^2)$$

I will adopt the convention of using a tilde to refer to random variables, so that \tilde{d} refers to the random dividend, while d_t refers to its realization at time t .

There are two types of agents: rational traders (denoted r) noise traders (denoted n). Rational traders have rational expectations about the economy, while noise traders believe

$$\tilde{d}_t \sim_n N(\bar{d} + \rho_t, \sigma_d^2)$$

where

$$\tilde{\rho}_t \sim N(\bar{\rho}, \sigma_\rho^2)$$

and the $\tilde{\rho}_t$ are i.i.d. with $\tilde{\rho}_t$ and \tilde{d}_t independent. All agents within a given type are identical.

I assume that there exists a continuum of both agent types with total measure \tilde{T} .⁷ Because total participation must be less than or equal to \tilde{T} , I assume that $\tilde{T} = \infty$, so that participation is potentially unbounded.

Each agent is born with wealth W_0 from which he or she may invest. All agents live for two periods, so having no further use for financial assets when old, they will liquidate their holdings and consume all of their wealth. Agents act so as to maximize a mean variance utility function of their wealth when they die:

$$U(W) = aE_0[W] - \frac{1}{2}a^2V_0[W] \quad (2.1)$$

Agents make two decisions: first, to enter the stock market, and second, how much to invest upon entering. Entering the stock market has a cost c , and if an agent pays that cost, he or she has access to the stock market for the remainder of his or her life. The entry decision and the investment decision are made in sequence, so the agents know only the distribution of prices when making their entry decision, while they know the price realization when placing their demand. The noise traders do not observe their signal until after they enter the market, and their signal will be fully

⁷This model is very similar to one in which noise traders have diverse signals. Because there is a continuum of traders, I can appeal to the law of large numbers to show that the idiosyncratic component of the noise trader signals will not affect the stock price. Diverse signals would have an effect on expected utility different from but similar to that of identical signals, but since there would be no effect on prices, I exclude any idiosyncratic component of signals for expositional simplicity.

revealed in prices. The bond market can be entered freely.

I interpret $\tilde{\rho}$ to be an uninformative public news signal that noise traders believe to be informative. This signal is revealed in two parts: $\bar{\rho}$ is known to all and is constant to reflect that much business analysis does not change rapidly – no newscaster is going to seriously consider General Motors to be a growth stock. A second portion, $\tilde{\rho}_t - \bar{\rho}$ is revealed only after agents have chosen whether they will be market participants; this represents residual uncertainty in market sentiment that changes with high frequency. This residual market sentiment cannot be predicted and so it must remain unknown when agents make their participation decisions.

I could alternately interpret $\tilde{\rho}$ to be a hedging demand: If noise traders hold a random number of units of human capital $\tilde{h}_t \sim N(\bar{h}, \sigma_h^2)$ with payoff $\tilde{g}_t \sim N(0, \sigma_g^2)$, then if $\text{Cov}(g_t, d_t) = \zeta$ and we set $\rho_t = \zeta h_t$, noise traders with hedging demand are identical to noise traders receiving an uninformative signal. In fact, the hedging interpretation is just as good as the belief differences interpretation: I call the agents “noise traders” because they introduce noise into the price, not because of where the noise comes from.

There is a large difference between my noise traders and the assumption of an exogenous supply shock in the stock. My noise traders are risk averse: when an investment is risky, they invest less and so less noise enters the stock price. This is reasonable because noise comes from investors, and investors optimize. Therefore noise responds to market conditions. The alternative to noise traders, supply shocks, enter the market with fixed and exogenous magnitude. They introduce noise in a way that does not react to market conditions. This is untenable in any model lasting more than two period because it ignores whatever decision-making process led to the shock.

To put it another way, because prices are formed from investor interaction, noise in prices must respond to the same forces that shape investor interaction. So, the magnitude of noise shocks must be responsive to market conditions.

I interpret c to be a barrier of entry into the stock market. In surveys of American

consumers, 63% of households own financial assets outside of transaction accounts, and 45% of households own financial assets outside of transaction accounts and tax-deferred accounts. (Bergstresser and Poterba (2003)). This is incompatible with most models of stock market demand without barriers to entry because even the most risk averse household will have a small demand for any given asset. Therefore there must be some barrier to participation in the stock market. c is the barrier to entry. ⁸

It is very important that the realized value of $\tilde{\rho}_t$ is only learned after the entry decision is made and the cost of entry paid. This can be justified in at least two ways. First, if a particular rational investor decides to become active in a stock, that decision must be made before he realizes all available information about that stock. In this case, the only valuable information that can be learned in the market is the realization of $\tilde{\rho}_t$. Similarly, the noise trader, when learning about the market, must learn something he considers valuable – like the rational trader, he learns the realization of $\tilde{\rho}_t$. Second, there is a meaningful distinction between participating in a marketplace and placing an order in that market. If this were not the case, it would not be the case that so many households have zero holdings in the stock market. Those investors are simply do not respond to short term changes in the price of a given stock. Because the participation and demand decisions are separate, it makes sense to assume that there is a time lag between the two. Market sentiment will change during the time between the two decisions, so that investors cannot make forecasts without error. This error is represented by $\tilde{\rho}_t - \bar{\rho}$.

I also assume that $\bar{\rho} > 0$. This is a reflection of that fact that IPOs tend to come in cycles (Lowry and Schwert (2002)). Firms will wish to take advantage of any positive effect on price that noise traders have, and so they will choose to go public when $\bar{\rho}$ is positive. This assumption also makes the results of noise trader participation in section 2.2.3 more striking.

⁸One example of where such a cost could arise is from the need to educate oneself about stock return processes and institutional details before getting involved in the stock market. Attempting to understand a market, even at a basic level, represents a significant investment in time and effort.

Risk Aversion

It is somewhat unusual that I am pricing an individual stock with investors that are risk averse. Both empirical and theoretical considerations make this a viable assumption. The underpinning of the belief that investors are risk neutral toward individual stocks is based on the diversification intuition: if there are many stocks, then any given stock should make up only a small portion of an investor's portfolio, and so that investor should be relatively unconcerned about idiosyncratic risk. Both parts of the diversification intuition have recently come under attack.

First, noise trader risk is partly systematic – industries, such as technology, often rise and fall together. Anecdotally, the recent technology bubble is an example of systematic mispricing. More generally, studies such as Barberis et. al. (2003) and Barberis and Shleifer (2003) have shown that stocks within a given category can co-vary more than their cash flows would indicate, while stocks in different categories will co-vary less.

Second, the assumption of risk aversion flows naturally from the assumption of imperfect arbitrage. The noise trader will attempt to exploit his private valuation in the stock, either because he believes he has information that the market lacks, or because he has a hedging demand that requires large positions. In either case, the noise trader will be willing to hold an undiversified portfolio and so will be risk averse with respect to the idiosyncratic risk of the larger part of that portfolio. The rational trader, in contrast, sees the noise trader's extra demand as a source of mispricing, and so he will be willing to take a large position so as to exploit the mispricing. In fact, this is exactly an arbitrage argument: in order to exploit arbitrage opportunities, the rational trader must take large positions and so he will be sensitive to idiosyncratic risk.⁹

Empirically, many studies have shown that household investors do not hold di-

⁹If instead, one assumes that there are some noise traders in every stock, then all that is necessary to create undiversified portfolios is that the different noise sources have some systematic component. If, for example, noise was correlated across technology stocks, investors would hold undiversified portfolios with respect to that industry, and they would therefore take account of industry specific noise risk.

versified portfolios¹⁰. There are several proposed explanations for this, including a barrier to entry (as I assume) or a preference for familiarity. Simultaneously, the idea that idiosyncratic risk is not priced has come under increasing attack. Goyal and Santa-Clara (2003) and Goetzmann and Kumar (2003) show that individual stock returns vary with idiosyncratic risk and with individual investor diversification.

If the median investor holds only one or two stocks, as the studies previously mentioned suggest, and idiosyncratic risk is indeed priced, then an assumption of risk aversion on behalf of investors is entirely plausible. In that case, uncertainty over future re-sale options is a major determinant of market conditions and of investment opportunity sets.

Before continuing, I must define two important pieces of notation:

- T_t is the total participation in the stock market in period t .
- μ_t is the fraction of total participation that is comprised of noise traders. So $\mu_t T_t$ is the total number of noise traders in the marketplace.

I also want to explain in detail the time-line associated with the two period model:

- In period 0, agents decide whether they wish to access the stock market; they must pay a cost c in order to do so.
- After entry, the noise trader receives his private value ρ_t , and then all agents place demands and trade. The noise trader's value will be revealed in prices.
- In the second period, the stock pays its dividend, \tilde{d}_t and is liquidated. All investors receive mean-variance utility over their terminal wealth.

I will further explain and motivate the behavior of the agents in the next two sections.

¹⁰A sample of such studies would include Bertaut (1998), Goetzmann and Kumar (2003), Grinblatt and Keloharju (2001), Kelly (1995), and Odean (1999)

2.2.2 The Stock Price and Expected Utility

To calculate the stock price and expected utility, I will take participation as given. For ease of exposition, I define $V_{d,t+1}$ as the risk inherent in owning the stock. For example, if the stock pays a dividend next period before being liquidated, then $V_{d,t+1} = \sigma_d^2$.

Because the young agents have mean-variance preferences (2.1), they place the standard linear demands:

$$x_{r,t} = \frac{1}{a} \frac{\bar{d} + E p_{t+1} - R p_t}{V_{d,t+1}} \quad \text{and} \quad x_{n,t} = \frac{1}{a} \frac{\bar{d} + E p_{t+1} - R p_t}{V_{d,t+1}} + \frac{1}{a} \frac{\rho_t}{V_{d,t+1}} \quad (2.2)$$

Since the aggregate supply of the stock is S , equation (2.2) and market clearing¹¹

$$(1 - \mu_t) T_t x_{r,t} + \mu_t T_t x_{n,t} = S$$

imply ¹²

$$p_t = \frac{1}{R} \left(\bar{d} + E p_{t+1} + \mu_t \rho_t - a \frac{S}{T_t} V_{d,t+1} \right) \quad (2.3)$$

In the pricing equation, $\mu_t \rho_t$ represents the direct effect of noise trader demand on prices. It is also the part of prices that both agent types regard as mispricing. The rational trader sees this as the effect of misinformed noise trader beliefs. The noise trader, however, sees it as the result of rational traders not taking into account an informative signal. Both traders view this part of the price as an opportunity to exploit the beliefs of the other.

Because both traders view price and dividend variance in the same way, they agree

¹¹One complaint that is often raised against this type of model has to do with market clearing: I do not clear the bond market, and if I did, all noise in the price would vanish. This effect is not deep: if investors have a fixed amount of wealth and bonds are in zero net supply, all the wealth must go into the stock. The price is then fixed at the level of aggregate wealth. The problem is that clearing the bond market removes any aggregate margin of choice that investors might have. Restoring a choice, for example by allowing early consumption and a savings decision, restores noise to the stock price. I do not clear the bond market for analytical convenience, but it should be understood that it is quite possible (with any valid utility function; mean-variance utility does not make sense with intermediate consumption).

¹² $\bar{\rho}_t$ is i.i.d., so while the rational trader can infer ρ_t from prices, he does not need to do this because ρ_t carries no useful information.

on what the risk premium should be, and that value is $a\frac{S}{T_t}V_{d,t+1}$. The term represents the premium each agent would require if each agent held an equal fraction of the outstanding supply of stock.

To calculate the expected gain from entry, I can plug equations (2.2) and (2.3) into the utility function (2.1). The item of interest for participation, however, is the agents' expected utility before they enter and observe ρ_t . These expected gains from entry are

$$EU_{r,t} = \frac{1}{2} \frac{\left(-\mu_t \bar{\rho} + a\frac{S}{T_t}V_{d,t+1}\right)^2}{V_{d,t+1}} + \frac{1}{2} \frac{\mu_t^2 \sigma_\rho^2}{V_{d,t+1}} \quad (2.4a)$$

$$EU_{n,t} = \frac{1}{2} \frac{\left((1 - \mu_t) \bar{\rho} + a\frac{S}{T_t}V_{d,t+1}\right)^2}{V_{d,t+1}} + \frac{1}{2} \frac{(1 - \mu_t)^2 \sigma_\rho^2}{V_{d,t+1}} \quad (2.4b)$$

These expected gain formulas can be interpreted as functions of the investment opportunity set. Each investor faces an opportunity set comprised of two parts: both agents perceive the stock to be mispriced, and they both observe and require a risk premium in order to hold the stock. The first term is the squared value of the average opportunity – the $\bar{\rho}$ piece is the perceived average mispricing, while the $V_{d,t+1}$ piece is the risk premium. Because the two agents agree on variances, the risk premium they require is the same for both; it is the perceived mispricing that varies from agent to agent. The second term in the expected utility formula is the variance of the investment opportunity set – the variance of the perceived mispricing. Because investors trade more aggressively when they perceive more mispricing, and because their expected profits per trade go up with additional mispricing, the value investors place on additional opportunities increases in the total value of opportunities. This convex return to opportunities is reflected in the fact that expected utility increases in the variance of those opportunities, represented by σ_ρ^2 .

2.2.3 Stock Market Participation

Agents will choose to enter if the benefit of doing so outweighs the cost; if

$$EU_{i,t} \geq c$$

Therefore an equilibrium requires that no trader can realize expected utility greater than c from entering:

Definition 3 *Entry Equilibrium:* A pair (μ_t, T_t) is an entry equilibrium for time t , if, given $V_{d,t+1}$, one of the following sets of conditions is met:

- $\mu_t = 0, T_t > 0, EU_{r,t} = c, EU_{n,t} < c$
- $\mu_t = 1, T_t > 0, EU_{r,t} < c, EU_{n,t} = c$
- $\mu_t \in [0, 1], T_t > 0, EU_{r,t} = c, EU_{n,t} = c$

With these conditions, agents enter until they are indifferent between participation and remaining outside the market, subject to the constraint that the number of entrants of each type must be at least zero. Rents are earned by participants to exactly offset the barrier to entry – rational and noise traders have the same expected utility gain from entering, c , when evaluated under their own probability measures.

I also wish to define a notion of entry stability: an equilibrium is stable with respect to entry if, taking all future stock moments as given, entry in this period of a given agent will decrease the expected utility of other agents of that type:

Definition 4 *Entry Stability:* A given entry equilibrium has entry stability if and only if $\frac{\partial EU_r}{\partial(1-\mu_t)T_t} < 0$ and $\frac{\partial EU_n}{\partial\mu_t T_t} < 0$.

“Entry Stability” means that in the given time period, we are not at a knife edge with respect to entry. It also means that gains from trade are distributed in a manner consistent with bargaining or an auction: an agent sees his gains decrease when someone just like him enters the market.

Lemma 3 implies that all equilibria in my model will exhibit entry stability:

Lemma 3 *Entry Stability:* $\frac{\partial EU_r}{\partial(1-\mu_t)T_t} < 0$ and $\frac{\partial EU_n}{\partial\mu_t T_t} < 0$.

“Entry stability” is the most reasonable definition of stability in an participation model such as mine. What is relevant to entry stability is not the effect of changes in future participation decisions – investors take those as given – but instead changes in current participation. To be stable, it must be the case that entry “errors” are self correcting – that if too many traders enter the market, they must be better off if they had not entered. This means that expected utility decreases in entry by ones own type, exactly my definition of “entry stability”.

The next step is to find the equilibrium values of μ_t and T_t . Because in each period, potential entrants take future participation as given, I will solve for current participation as a function of the risks associated with holding the stock, $V_{d,t+1}$. This characterizes the system in a one-period setting; later, I will show how using a dynamic entry framework will allow for a much richer set of states. Remember that I have restricted my analysis to the case where $\bar{\rho} > 0$.

Lemma 4 (μ_t, T_t) is an entry equilibrium¹³ if:

The Corner Solution: If $c > C_1$, then $\mu_t = 1$ and $T_t = \frac{aS}{\sqrt{2c}}V_{d,t+1}$.

The Interior Solution: If $C_{min} < c < C_1$, then

$$\mu_t = \frac{1}{2} + \frac{1}{2} \frac{\bar{\rho}}{\sigma_\rho} \sqrt{\frac{8cV_{d,t+1}}{\sigma_\rho^2 + \bar{\rho}^2} - 1} \quad (2.5a)$$

$$T_t = T(\mu_t) = aS \frac{V_{d,t+1}}{\sigma_\rho^2 + \bar{\rho}^2} \frac{\bar{\rho}}{\mu_t - \frac{1}{2}} \quad (2.5b)$$

Otherwise, there is no equilibrium.

When $c < C_{min}$ there is no equilibrium because no amount of entry will ever reduce investors’ gains to c . This means that all potential investors want to enter, but because I have specified that there are an infinite number of potential investors, there is no specific number that creates an equilibrium. If I were to fix a maximum

¹³The values of C_1 and C_{min} are uninformative and can be found in the appendix.

population of potential entrants, then I would see an equilibrium for all values of c , but that equilibrium would be determined, not by the gain from entry, but by my choice exogenous population parameters. To avoid the illusion inherent in creating such an equilibrium, I state that there is unlimited potential entry and no equilibrium for low values of c .

Figure 2-1 shows graphically how μ_t and T_t change in response to $V_{d,t+1}$. Because $V_{d,t+1}$ is the only relevant future variable, this plot gives a view of the cross section of market composition as the future changes.

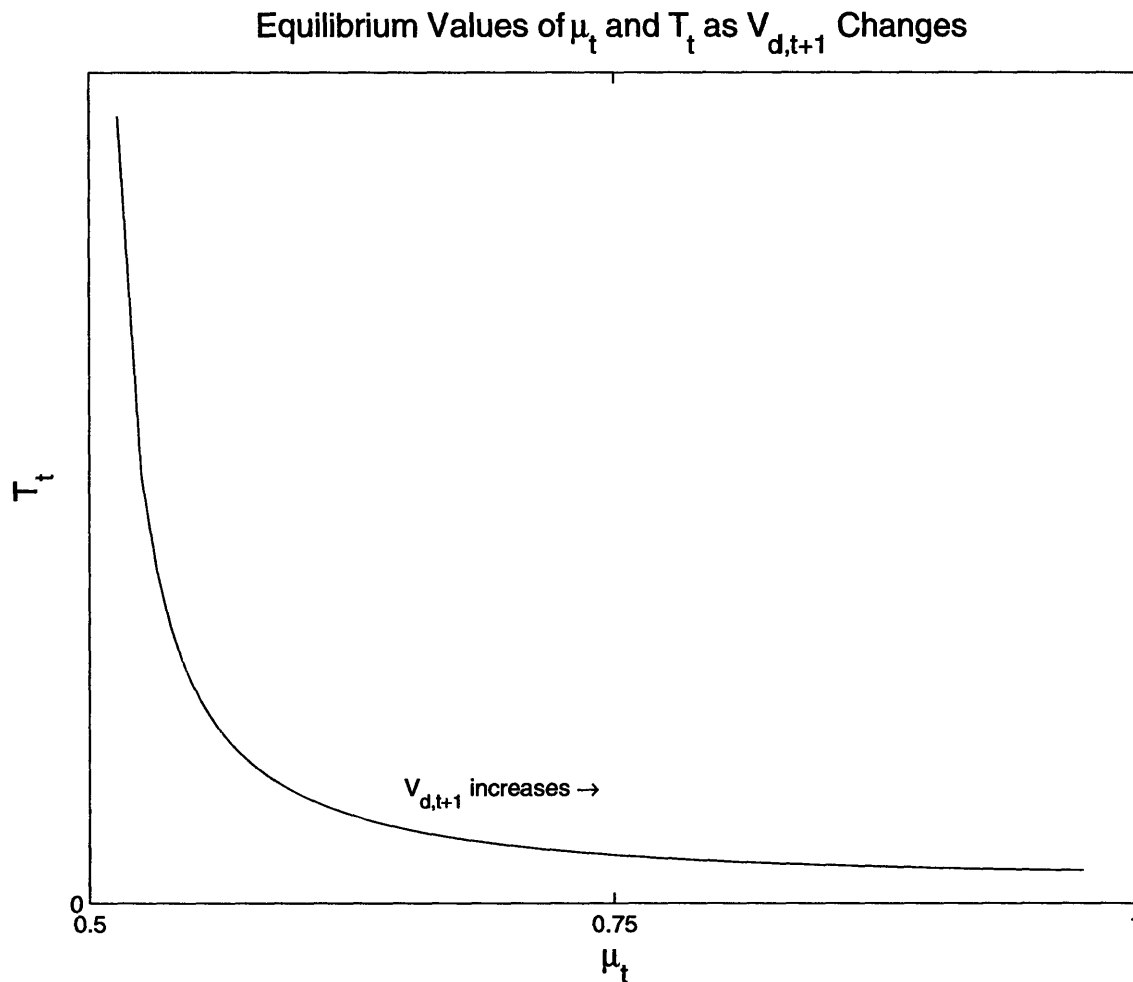


Figure 2-1: Lemma 4 describes the cross section of market participation as future risk changes.

Total Participation

Total participation is determined endogenously so as to characterize the risk premium. In fact, T_t appears in the pricing and expected utility equations (2.3 and 2.4) only as part of the risk premium, $a\frac{S}{T_t}V_{d,t+1}$. Moreover, T is the only endogenous part of the risk premium, so determining total participation endogenously is the same as determining the risk premium through entry conditions.

The expression for $T(\mu_t)$ in equation 2.5 is obtained by setting $EU_{r,t}$ and $EU_{n,t}$ equal to one another. I can therefore use the expression for T_t without substituting μ_t to do a kind of comparative static: I will look at the effect of market composition, μ_t on participation, T_t . To make this rigorous, it should be understood that I am looking across equilibria, and that in the background I am changing c so as to vary μ_t . Moreover, by allowing μ_t to change rather than explicitly changing c , I am restricting what follows to interior solutions.

To see how market composition affects the risk premium and total participation, I re-write equation 2.5 as

$$a\frac{S}{T_t}V_{d,t+1} = \left(\mu_t - \frac{1}{2}\right) \frac{\bar{\rho}^2 + \sigma_\rho^2}{\bar{\rho}}$$

where the left hand side is the risk premium determined by trade in the market, while the right hand side is the resulting premium when total participation is determined endogenously.

The key aspect to setting the risk premium through participation is that when $\bar{\rho} > 0$, the rational trader can only collect the premium for holding the stock by going long, while he can only exploit the mispricing by going short.¹⁴ The noise trader, on the other hand, can both collect the premium and exploit the perceived mispricing by going long. Participation will adjust the risk premium so that the total payoffs to the rational and noise traders will have the same value.

¹⁴In equilibrium, the rational trader will have a position that is on average long if $\bar{\rho}^2 < \sigma_\rho^2(2\mu - 1)$. This simply means that the rational trader will exploit the mispricing rather than take the risk premium if the mispricing is large enough.

The first term in the endogenous risk premium is $\mu - \frac{1}{2}$. This implies that the risk premium increases as μ increases and that the risk premium is zero when $\mu = \frac{1}{2}$.¹⁵ When $\mu = \frac{1}{2}$, both traders perceive same degree of mispricing, so to equate payoffs, the rational investor must be, on average, as short and the noise investor is long. This is only possible if the risk premium is zero and the total number of participants is very large.

As μ_t increases, the rational trader begins to perceive that the stock is increasingly mispriced. This would increase the opportunity set of the rational agent, but because participation is endogenous, the total value to entry must remain the same. The key is that because $\bar{\rho} > 0$, the rational trader sees the risk premium and the mispricing working against one another: to exploit mispricing, he must go short, but to obtain the premium from holding the stock, he must go long. The larger is μ , the more mispricing he sees, and so the risk premium must increase commensurably so as to keep his expected utility gains constant. Therefore the risk premium must increase in μ .

The second term in the risk premium is $\bar{\rho} + \frac{\sigma_{\rho}^2}{\bar{\rho}}$. The direct effect of $\bar{\rho}$ works analogously to an increase in μ – it increases the mispricing seen by the rational trader. The second effect, from $\frac{\sigma_{\rho}^2}{\bar{\rho}}$, is simply the relative value of the range in $\tilde{\rho}$ over its mean. When $\mu > \frac{1}{2}$, the rational trader perceives more variable mispricing than the noise trader, and that variation is parameterized by σ_{ρ}^2 . As that variation increases, so does the risk premium; again, the argument is similar to that for μ_t .

Pricing

As a result of endogenous participation, I can use T_t to re-write the pricing function, equation 2.3, as

$$E p_t = \frac{1}{R} \left(\bar{d} + \mu_t \bar{\rho} - a \frac{S}{T_t} V_{d,t+1} \right) = \frac{1}{R} \left(\bar{d} - \mu_t \frac{\sigma_{\rho}^2}{\bar{\rho}} + \frac{1}{2} \bar{\rho} + \frac{\sigma_{\rho}^2}{2\bar{\rho}} \right) \quad (2.6)$$

¹⁵Because T_t is positive, and because therefore the risk premium must be positive, equation 2.5 implies that $\bar{\rho} > 0 \Leftrightarrow \mu_t > \frac{1}{2}$.

and I can continue to look across interior equilibria.

The most important result from this equation is that expected price *decreases* as the fraction of noise traders increases, even when the noise traders are optimistic. This effect has nothing to do with any noise trader risk in tomorrow's price – remember that I've kept $V_{d,t+1}$ constant – but instead is an effect of participation. When the fraction of noise traders in the market, μ_t , increases, they have a positive direct effect of prices through their increased demand, but they have a negative indirect effect through total participation. When there are more noise traders in the market, that market will have much less aggregate participation and a higher risk premium. The market's ability to absorb risk is so diminished that the expected price of the asset falls even in excess of direct demand of noise traders for the stock. Thus, increasing relative noise trader participation lowers expected prices *regardless of noise trader risk* next period.

The increase in the risk premium and the fall in total participation dominate the increase in direct demand from the noise traders because participation responds to both the variance and average level of mispricing. Conversely, the direct demand effect is simply the average level of mispricing. Because the variance in mispricing exerts an effect on participation and the risk premium over and above the effect of average mispricing, the participation effect must dominate the direct demand effect. The result is that μ decreases the expected price.

The second result is that when noise traders become more optimistic – $\bar{\rho}$ increases – then the expected price of the stock increases. While it is true that increasing $\bar{\rho}$ will have a similar effect on participation as increasing μ , the effect is not nearly as strong because $\bar{\rho}$ affects only the average level of mispricing while μ affects both the average level of mispricing and its variance. Because both the average and variable level of mispricing affect agents' expected utility, the participation effect of μ is stronger than that for $\bar{\rho}$. As a result, the direct effect of increasing $\bar{\rho}$ – increasing the noise traders demand for stocks – can outweigh the indirect participation effect. This is not the case for μ .

A second way to understand the noise traders effect on the risk premium is to ask what happens to total participation and to prices if one changes the cost of entry faced by noise traders. So if an entry equilibrium is of the form

$$EU_{r,t} = c - \Delta \text{ and } EU_{n,t} = c + \Delta$$

then equation 2.5 becomes

$$T(\mu) = aS \frac{\bar{\rho} V_{d,t+1}}{(\sigma_\rho^2 + \bar{\rho}^2) (\mu_t - \frac{1}{2}) - 2\Delta V_{d,t+1}}$$

and

$$Ep_t = \frac{1}{R} \left(\bar{d} - \mu_t \frac{\sigma_\rho^2}{\bar{\rho}} + \frac{1}{2} \bar{\rho} + \frac{\sigma_\rho^2}{2\bar{\rho}} + 2\Delta \frac{V_{d,t+1}}{\bar{\rho}} \right)$$

where $\frac{\partial \mu_t}{\partial \Delta} < 0$. So $\frac{\partial Ep_t}{\partial \Delta} > 0$, meaning that if the firm could raise the relative cost of noise trader entry, doing so would increase the expected value of the stock. While the direct demand effect of noise traders on the price would decline because they form a smaller fraction of the population, the indirect and positive effect of participation would be larger.

The situation where $c < C_{min}$ does provide a useful way to compare my model with that of De Long et al (1990). They examine the effect of a fixed market composition on prices and obtain some results that are different from mine. For example, in their model prices can increase or decrease as a result of noise trader participation. This is because they are changing the composition directly while I am looking across entry equilibria. So when they change μ , total participation remains fixed which is not the case in this model. However, if I were to fix the maximum population size and set c close to zero, then all agents would want to enter. If I changed the fraction of noise traders in the population as a whole, I would also change the number of noise traders in the entry equilibrium. The reason for this is that both agent types would be at a corner – any outside agents would want to enter if they existed. The comparative statics in the DSSW model are analogous to mine in the corner solution where all

agents enter, but in that case there is no meaningful margin of participation.

Lastly, notice that the aggregate supply of the stock, S , and the risk aversion of the agents, a , do not enter the equilibrium stock price. This is because allowing total participation to be set endogenously means that the risk premium is set endogenously. S and a affect the total amount of risk in the market and how much investors must be compensated for holding it, but they both have their effect only because a market has a given depth.¹⁶ Those variables do affect total participation – they affect how many people are needed to achieve a given risk premium – but they do not affect the endogenously determined premium itself. Participation totally undoes the effects of aggregate supply – if supply doubles, so will participation. As a result, S and a have no effect on the stock price or on expected returns.

Market Composition

To examine market composition, μ_t , I will use a comparative static taken across equilibria: $\frac{\partial \mu_t}{\partial V_{d,t+1}}$. Because I am looking at changes in μ , I will examine only interior solutions.

If T_t captures the risk premium, then μ characterizes the level and variation of mispricing. The important part of equation 2.5,

$$\mu_t = \frac{1}{2} + \frac{1}{2} \frac{\bar{\rho}}{\sigma_\rho} \sqrt{\frac{8cV_{d,t+1}}{\sigma_\rho^2 + \bar{\rho}^2} - 1}$$

for μ is that it increases when the total risk of holding the stock increases: $\frac{\partial \mu_t}{\partial V_{d,t+1}} > 0$. The second important piece of equation 2.5 is that if $\bar{\rho} > 0$, then $\mu_t > \frac{1}{2}$; this latter fact just reflects that whoever places a high average value on the stock will make up more than half of the total market participants.

To understand why μ_t increases in $V_{d,t+1}$, I must go back to the equations that

¹⁶There are other models that predict that markets can have multiple endogenous levels of depth, such as those by Allen and Gale (1994), Cespa (2002), and Pagano (1989). My model is different because of the way I involve noise traders and because of my focus on market composition.

specify expected utility (2.4) to show that

$$\mu > \frac{1}{2} \Leftrightarrow \frac{\partial EU_{n,t}}{\partial V_{d,t+1}} > \frac{\partial EU_{r,t}}{\partial V_{d,t+1}}$$

This means that if the noise trader makes up more than half the population, then when the variance of the stock increases, he loses less than the rational trader. This happens because when $\mu > \frac{1}{2}$, the noise trader perceives there to be less mispricing than the rational trader. Because the value of the investment opportunity set comes from both perceived mispricing and the risk premium, when the noise trader perceives less mispricing, the risk premium makes up a larger share of the value of his opportunity set. So when the risk premium increases, the noise trader will gain proportionately more utility. Increasing the risk of holding the stock then has two effects: first, it reduces everyone's value to participation equally – this results from the fact that both traders perceive risk in the same way. Second, increasing risk increases the risk premium which benefits the noise trader disproportionately. As a result, the direct effect of risk is to drive both agent types out of the market, but the increase in the risk premium brings some of the noise traders back in. So increasing risk increases the noise trader fraction of participation.

Looking again at the noise trader's utility,

$$EU_{n,t} = \frac{1}{2} \frac{\left((1 - \mu_t) \bar{\rho} + a \frac{S}{T_t} V_{d,t+1} \right)^2}{V_{d,t+1}} + \frac{1}{2} \frac{(1 - \mu_t)^2 \sigma_\rho^2}{V_{d,t+1}}$$

the important term is $\left((1 - \mu_t) \bar{\rho} + a \frac{S}{T_t} V_{d,t+1} \right)^2$, so the marginal effect of a small increase in the risk premium – $a \frac{S}{T_t} V_{d,t+1}$ – is $2(1 - \mu_t) \bar{\rho} + 2a \frac{S}{T_t} V_{d,t+1}$. Compare this to the rational trader's utility

$$EU_{r,t} = \frac{1}{2} \frac{\left(-\mu_t \bar{\rho} + a \frac{S}{T_t} V_{d,t+1} \right)^2}{V_{d,t+1}} + \frac{1}{2} \frac{(1 - \mu_t)^2 \sigma_\rho^2}{V_{d,t+1}}$$

where the marginal effect of a small increase in the risk premium is $-2\mu_t \bar{\rho} + 2a \frac{S}{T_t} V_{d,t+1}$.

Because the noise trader is optimistic ($\bar{\rho} > 0$) he gains more from the risk premium in expected utility terms than the rational trader. Because risk leads to a premium in the stock's price, the noise trader is more willing to tolerate that risk when making his entry decision.

2.2.4 The Three Period Aftermarket

I will now extend the two period model described in the previous three sections to a three period model. I do this because I want to allow agents to make decisions more dynamically and to illustrate the nature of the market participation externality more clearly. A three period model enables me to do these things because it provides two trading periods in which the risk in the first is characterized by participation in the second. Because agents do not take this into account when making entry decisions, the result is a participation externality.

The three period economy will proceed as follows:

- At time 0, all traders decide if they will pay a cost c to have access to the stock market for the duration of the economy. Noise traders who choose to enter receive a private signal ρ_1 which reflects their beliefs about the dividend \tilde{d}_1 . Agents then trade S shares among themselves.¹⁷
- At time 1, the dividend d_1 is realized but not paid, and noise traders receive their private signal ρ_2 which reflects their belief about the dividend d_2 . Agents again trade among themselves, but there is no additional entry.
- At time 2, the stock pays a terminal dividend, $d_1 + d_2$ and the economy ends. All traders receive utility over their terminal wealth.

The economic variables and agent preferences will be as described in section 2.2.1, with the exception that for simplicity I will normalize the risk free rate to zero. Moreover, I will use the same definition of an entry equilibrium (Definition 3).

¹⁷No trader is endowed with any shares of the stock. Instead, there is some seller who exogenously sells S shares at whatever price the market will bear.

In this economy, both the rational and noise traders fully understand both the economic structure and the distribution of prices. As a result, the noise trader makes his entry decision knowing that he will receive two signals and believing both signals to be informative. Simultaneously, he know the variability of the stock price in the second period and accounts for it as a price taker in his entry decision. The noise trader does not act strategically: because he is a price taker he does not account for the fact that he is causing the pricing variability. Similarly, the rational trader does not take account of the fact that by entering he reduces volatility. This lack of strategic action is one manifestation of the participation externality: each agent makes his or her entry decisions without accounting for his or her effect on other agents or on the price.

Because agents have mean-variance utility, they will place linear demands, and prices can be solved recursively,

$$p_0 = 2\bar{d} + \mu\rho_0 + \mu\bar{\rho} - a\frac{S}{T}(\mu^2\sigma_\rho^2 + 2\sigma_d^2) \quad (2.7a)$$

$$p_1 = (\bar{d} + d_0) + \mu\rho_1 - a\frac{S}{T}\sigma_d^2 \quad (2.7b)$$

In each pricing equation, the first term represents the expected value of future dividends, while the second term represents the direct effect of noise traders on prices. The term $\mu\bar{\rho}$ in the first equation is the effect of expected noise trader demand in the second period. Finally, the last term is the risk premium represents compensation for the risk investors bear for holding the stock; it is higher in period 0 because the stock price in period 1 is subject to noise trader risk. Notice as well that prices fully reveal the value of the noise traders' signal.

Using the above pricing equations and the fact that investors have mean-variance

utility, I can calculate the expected gain from entry to the two investor types:

$$\begin{aligned}
EU_{r,t} &= \frac{1}{2} \frac{(-\mu\bar{\rho} + a\frac{S}{T}(\mu^2\sigma_\rho^2 + \sigma_d^2))^2}{\mu^2\sigma_\rho^2 + \sigma_d^2} + \frac{1}{2} \frac{\mu^2\sigma_\rho^2}{\mu^2\sigma_\rho^2 + \sigma_d^2} + \frac{1}{2} \frac{(-\mu\bar{\rho} + a\frac{S}{T}\sigma_d^2)^2}{\sigma_d^2} + \frac{1}{2} \frac{\mu^2\sigma_\rho^2}{\sigma_d^2} \\
EU_{n,t} &= \frac{1}{2} \frac{((1-\mu)\bar{\rho} + a\frac{S}{T}(\mu^2\sigma_\rho^2 + \sigma_d^2))^2}{\mu^2\sigma_\rho^2 + \sigma_d^2} + \frac{1}{2} \frac{(1-\mu)^2\sigma_\rho^2}{\mu^2\sigma_\rho^2 + \sigma_d^2} \\
&\quad + \frac{1}{2} \frac{((1-\mu)\bar{\rho} + a\frac{S}{T}\sigma_d^2)^2}{\sigma_d^2} + \frac{1}{2} \frac{(1-\mu)^2\sigma_\rho^2}{\sigma_d^2}
\end{aligned}$$

This is quite comparable to equation 2.4: each investor receives expected utility each period from the expected and variable parts of his opportunity set. The difference is that now agents must act dynamically: the rational trader will take account of both his ability to exploit the noise trader and the fact that the noise trader increases the risk of holding the stock.

These expected utilities share the property of “entry stability” with their two period analogs:

Lemma 5 *Entry Stability:* $\frac{\partial EU_r}{\partial(1-\mu_t)T_t} < 0$ and $\frac{\partial EU_n}{\partial\mu_t T_t} < 0$ for $\mu \in [\frac{1}{2}, 1]$.

This means that no equilibria that I find will be on a knife edge with respect to entry – entry always decreases the expected utility of that agent’s type.

Much like in section 2.2.3 I can solve for entry equilibria. The different values of C are analogous to their counterparts in the two period model (lemma 4) and can be found in the appendix.

Lemma 6 (μ, T) is an entry equilibrium if:

The Corner Solution: If $c > C_1$, then $\mu = 1$ and $T = \frac{aS}{\sqrt{2c}}\sqrt{3\sigma_d^2 + \sigma_\rho^2}$.

The Interior Solution¹⁸:

- If $C_{min} < c < C_1$, then there exists a $\mu^H \in [\frac{1}{2}, 1]$ that is an equilibrium.
- If $C_{min} < c < C_{\frac{1}{2}}$, then there exists a $\mu^L \in [\frac{1}{2}, 1]$ that is an equilibrium.

¹⁸The values of μ^H and μ^L are the two roots between $\frac{1}{2}$ and 1 of an eighth degree polynomial. That polynomial can be found in the appendix.

- If both μ^H and μ^L exist, then $\mu^H > \mu^L$.
- If either μ^H or μ^L exists, then

$$T = T(\mu) = 2aS \frac{1}{\bar{\rho}^2 + \sigma_\rho^2} \frac{\bar{\rho}}{\mu - \frac{1}{2}} \frac{\sigma_d^2 (\mu^2 \sigma_\rho^2 + \sigma_d^2)}{\mu^2 \sigma_\rho^2 + 2\sigma_d^2}$$

Otherwise, there is no equilibrium¹⁹.

Much of this result is analogous lemma 4: total participation is determined endogenously so as to characterize the risk premium, while market composition characterizes the level of mispricing. The difference is that now there are two equilibria – two levels of mispricing. Remember that in the two period model, the equilibrium value of μ_t increases with the risk of holding the stock. In the three period model, more noise traders makes the stock in period 1 more risky. While this effect is not there is the final trading period, because the economy ends the next day, one period is enough to create multiple equilibria. Additional noise traders increase the relative payoff of noise traders in the first period and as a result there are two equilibrium values for investor composition, μ .

While 6 shows the existence of two equilibria, I must make sure that they share the same properties as their previous analogs – that expected price decreases with μ . Furthermore, I want to show that not only does total participation move inversely to μ , but so do rational and noise trader participation individually:

Lemma 7 Pricing: *If both μ^H and μ^L exist, then $Ep(\mu^L) > Ep(\mu^H)$.*

Participation: *If both μ^H and μ^L exist, then $\mu^H T^H < \mu^L T^L$, $(1 - \mu^H) T^H < (1 - \mu^L) T^L$, and $T^H < T^L$.*

The intuition for these results is much the same as in the two period model: increasing the fraction of noise traders has such a severe effect on participation and the risk premium that the stock price declines. In fact, increasing the fraction of noise

¹⁹If $c < C_{min}$ there is no equilibrium because no amount of entry can reduce expected utility below c .

traders causes such a precipitous drop in participation that the actual level of noise trader entry declines: when $\mu = \mu^H$, there are both fewer noise and fewer rational traders.

2.2.5 Discussion

In the three period model, there can arise endogenously two different sets of market participants and two different stock market behaviors. These effects are the result of a market participation externality – investors do not take account of their own entry on the condition of the marketplace – and it demonstrates the difficulty in forming an public market.

In the previous four sections, I have described the effects of endogenizing participation, such as the fact that increasing the presence of optimistic noise traders lowers the expected stock price. This result different from those in De Long et. al (1990) – in their model noise traders have an ambiguous effect on the price due to the direct demand versus risk tradeoff. In my model, participation means that increasing the fraction of noise traders has a negative effect on the price regardless of noise trader risk next period.

The multiple equilibrium result is important because it means that investor types may tend to be found together endogenously. Because investors determine prices, this means that firm prices will show characteristics by category that are not reflected in their fundamentals. In one way this can be seen as endogenizing the investment “style” literature.²⁰ But it is also a challenge to the firm: if the firm wishes to maximize its value, how does it select the investors to do that? Lastly, I make clear the difference between participation and ownership. Participation is really about potential: does an agent have a demand curve, and does he or she exert market impact in the short term? Ownership is the result of the participation decision but participation is always much wider than ownership. In fact, as my modelling shows, it can be participation rather than ownership that has the larger impact on prices.

²⁰Please see Barberis and Shleifer (2003) and related papers for an overview of this literature.

The second result is that endogenizing participation decisions can have unexpected effects – a firm with relatively more noise trader participation will not have a higher price, even if those noise traders are extremely optimistic. The effects of heterogeneous agents, then, are not as straightforward as one might expect.

Finally, I want to make clear what my model says about fundamental value. If fundamental value is simply the expected payoffs to owning a stock discounted by some stochastic discount factor, then in my model a stock has many possible fundamental values. Because fundamental value is market specific – who participates in a market determines the stochastic discount factor – it only makes sense to discuss the notion of a fundamental value within the context of a given market. In fact, because different investors have different notions of the fundamental value of an asset, this value is endogenously determined even when everything about the firm is taken as given.

2.3 The IPO Allocation

Any firm that wishes to go public must undertake to create a secondary market for its securities. While this is not the only example of the creation of public trading markets, it is the most widely repeated. As a result, IPOs are an excellent place to start in studying the creation of trading markets. So I will place my question “How does one create a public trading market?” in an IPO framework.

I will assume that the firm going public has a fixed underlying technology, but that does not mean that the fundamental value of that technology is fixed. Because the firm, and by extension the firm’s technology, will be traded on a public market, the fundamental value of the firm is determined by the conditions under which it is traded. By creating a high quality secondary market, the firm is increasing its own fundamental value. As a result, it has the opportunity to increase the proceeds from the sales of claims on its underlying technology. In a sense, when a firm creates a public trading market in claims on itself, it creates its own fundamental value.

If the firm underprices, it is underpricing relative to the aftermarket, not to some exogenously given fundamental value.

I want to show how a firm can create a desirable public trading market, with the focus on the market itself. I will not create an optimal mechanism; instead, I will show, in way that does not depend on many specific institutional details, how a firm can go about marketing itself.²¹ Legal and institutional details vary so much from country to country, that any optimal mechanism must be country specific. I will demonstrate one method by which the problem of market formation can be overcome, and in doing so I will explain some of the puzzles in the IPO literature.

It is important to note that in order to make the market composition question meaningful, the firm or its investment bank must not be able to tell rational and noise traders apart *a priori*. Instead, it must use the methods available in the allocation procedure to induce the different agents to act to the firm's advantage. I also do not want to equate exactly "rational" with "institution" and "noise" with "retail". While institutions are undoubtedly more savvy than most retail investors, they are also more likely to have hedging motives for trade – institutions have strategic objectives, agency problems, and the like. The distinction between noise and rational trades in my model is that noise traders have an option on a private value. This is more likely to apply to individuals than to institutions, and I will return later to evidence that institutions do in fact act more like the rational traders in my model.

2.3.1 The Allocation Procedure and the Pre-Market

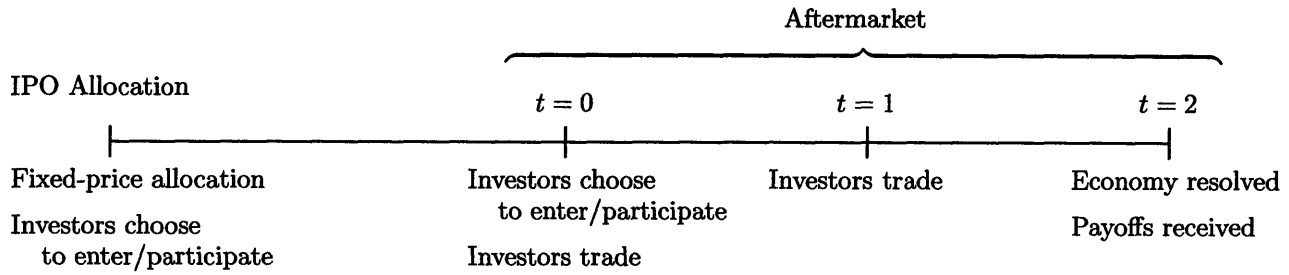
I assume that the allocation process proceeds²² as follows:

- The issuing firm announces that it will sell S shares of stock to investors. After the allocation these shares will be tradable in the three period aftermarket. The

²¹For theoretical and empirical discussions of optimal mechanisms in IPOs, please see Benveniste and Busaba (1997), Benveniste and Spindt (1989), Benveniste and Wilhelm (1990), Biais and Faugeron-Crouzet (2002), Chemmanur and Liu (2002), Derrien and Womack (2003), Spatt and Srivastava (1991), and Stoughton and Zechner (1998).

²²My allocation process is a fixed price allocation resembling that of Rock (1986) in many respects.

Figure 2-2: The Timeline. The dividend is paid at the beginning of periods 1 and 2.



company further announces the price p_a at which it is willing to transact.

- Investors can choose to pay a cost c_a to learn about the company and enter the allocation process.²³ In learning about the company, all traders learn the private valuation of the noise traders, ρ_a . This value of ρ_a represents the noise traders' incorrect beliefs about the stock's "fundamentals" and will perform the same role as $\tilde{\rho}_t$ in the aftermarket.
- At the same time that potential investors make their entry decision, the firm can decide with withdraw its IPO. If it does so, investors are refunded the cost c_a and the firm receives no proceeds. The simultaneous actions of the firm and investors reflects the fact that the firm can abort the pre-market process if it determines that not enough investors will eventually subscribe to the issue.
- All investors that paid the cost of entry may place a bid. This bid is a quantity and must be either zero or δ .²⁴

²³I interpret c_a as a barrier to entry into IPOs, much as I interpret c as a barrier to entry in the aftermarket. My preferred interpretation is that c_a is the cost of becoming knowledgeable about the IPO process and the companies involved. However, it could equally be interpreted as the cost of a relationship necessary to gain access to IPOs.

²⁴Imagine that investors can place any demand. Then, if the total demand at a given p_a is greater than the number of shares outstanding, there will be rationing in the allocation. Because all valuations and procedures are public information, all investors know how the rationing system will allocate shares, so all investors will scale up their demands to the degree they know shares will be rationed. This requires a tightening of the rationing rule, which forces investors to scale up their

- The firm allocates on an equal basis whatever shares it will sell: if it is selling a total of S shares and receives a total of δZ in bids, then it allocates δ shares to each bidder with probability $\min\left(1, \frac{S}{\delta Z}\right)$. If the firm is unable to allocate all of its shares, then it pays a cost M .²⁵
- Agents receive their shares and enter the aftermarket. If an investor participates in the allocation, he or she will also participate in the aftermarket. This means that c_a is the cost in excess of aftermarket utility gains that investors must pay. The firm will later use the fact that participants in the allocation also participate in the aftermarket to constrain the aftermarket equilibrium.

The first three stages are meant to represent the pre-market for an IPO: the firm announces what it intends to do, and then it and investors simultaneously decide whether to proceed. In reality, the firm goes on a “road show” where it solicits investor “indications of interest”.²⁶ If firm believes that investor interest is not sufficient, it can abort the road show and cancel its intended allocation. Because the investor’s decision to participate and the firm’s decision to continue the IPO are jointly determined, I model them as being simultaneous decisions.

I have also assumed that all traders and the firm understand the allocation system and all the variable distributions. Again, this is something that could be accomplished through a book-building procedure or other optimal mechanism. Moreover, there is no actionable information asymmetry in my model.²⁷ The noise trader’s signal is revealed to him immediately before placing his bids. While the signal is not revealed in his bids, no investor would change their action at that point if they knew the

demands even further, etc. There will be an equilibrium only if investors have a maximum possible demand, in this case δ .

²⁵I will assume that M is in fact infinite. This makes placing all of its shares a constraint that the firm is under.

²⁶Please see Jenkinson and Ljungqvist (2001) for a discussion of different institutional patterns, particularly across countries.

²⁷A second assumption about ρ_a is that it is independent of the realizations of ρ_t in the aftermarket. I make this assumption to avoid creating a situation with asymmetric information and because the signal is supposed to represent a temporary sentiment effect. It is $\bar{\rho}$ that represents the more permanent sentiment effect, and $\bar{\rho}$ remains the same in the allocation and in the aftermarket.

content of the signal. Therefore I could equally well model ρ_a as a public signal that only the noise traders believe to be informative.

It is important, as in the aftermarket, that the noise trader receives his signal only after he has made his entry decision and the allocation is irreversible. All agents know \bar{p} , but $\rho_a - \bar{p}$ is not known to anyone until after the noise traders investigate the stock. This can represent a staggered information release process: the pre-market was used to discover \bar{p} , but it was unable to discover all relevant information. In that understanding, $\rho_a - \bar{p}$ is simply the undiscovered content to the noise trader's valuation. Alternately, it could be that as the day of the issue approaches, the noise traders change their mind about the stock. Either way, there is some residual information contained in their signal that has not been revealed to anyone. However, because it is only revealed as noise traders enter the allocation, there is no asymmetric information – it is simply that the information is revealed to all after the allocation is already underway.²⁸

In a more complex model, I would need to include actions that the firm takes during the first few weeks or months after an IPO. These activities include stabilization, and market making activities among others. There is an established literature on the aftermarket activities of underwriters that concludes both that market making activities are important to underwriters and that price stabilization is common and integral to the underwriting process.²⁹ I exclude these facts from my model in order to simplify the exposition, but that does not mean that the facts are unimportant. In fact, the ability to stabilize the aftermarket could only help the firm in choosing its desired investor clientele because it would add new degrees of potential action. Again, an optimal IPO allocation mechanism would undoubtedly take direct aftermarket manipulation into account, but finding such a mechanism is not my goal.

²⁸ \bar{p} could also be interpreted as part of the IPO hot/cold market cycle. Besides the survey articles, Lowry and Schwert (2002) provide a discussion of the empirical basis for hot and cold IPO markets. In a hot IPO market, \bar{p}_a would be very high, increasing what a firm could ask for its shares. The underpricing would still be present, however.

²⁹Examples of this literature include Aggarwal(2000), Benveniste et. al. (1996), and Ellis et. al (2000).

Instead, I wish to show how the investor participation externality works and how a firm can use it to find a desirable ownership set.

2.3.2 The Investor's Problem

Because investors can accurately predict the entry behavior of other agents, once the firm announces p_a and S , all agents know the value of μ that will prevail in the aftermarket. In equilibrium, it will be the case that the firm never has to withdraw an IPO, so I will assume in this section that investors know this and act accordingly. Knowledge of the allocation process and of the strategies of other agents allows investors to calculate the expected price in the aftermarket and the expected difference between that price and the allocation asking price. I define underpricing as

$$\pi = \mathbb{E}p_0 - p_a \quad (2.8)$$

which corresponds to the expected value of the first trading day's return.³⁰ A second analytically useful definition of underpricing is the difference between the allocation price and a rational investor's valuation of one share from a portfolio of δ shares.

$$\chi = \mathbb{E}p_0 - \frac{a\delta}{2}Vp_0 - p_a \quad (2.9)$$

χ will be useful in simplifying the description of the allocation equilibrium, but it should be understood that π is the empirically observable variable.

It next remains to describe the behavior of the noise traders after they enter the allocation process: when do they exercise their option? Because the noise traders have mean-variance utility, their valuation of δ shares of stock offered at p_a is $\delta\mathbb{E}p_0 + \delta\rho_a - \frac{a\delta^2}{2}Vp_0 - \delta p_a$. Because the noise traders will bid whenever this value is greater

³⁰I am not stating return as a percentage because of the mean-variance and normal distribution assumptions in my setup. It would change nothing to define return as $\frac{\pi}{p_0}$.

than 0, (there is no short selling) they will bid whenever

$$\chi + \rho_a \geq 0$$

To simplify the exposition, I want to define two pieces of notation:

- $\Phi_\chi = \text{Prob}(\rho_a \leq -\chi)$. This is the probability that the noise trader does not bid on the asset; it is also the cumulative density function for ρ_a .
- ϕ_χ is the probability density function for ρ_a evaluated at $-\chi$.

I will now define the notion of an entry equilibrium in the allocation process. I will use $EU_{i,a}$ to mean the expected gain of an agent of type i obtained by entering the allocation.

Definition 5 Allocation Entry Equilibrium: A pair (μ_a, T_a) is an entry equilibrium in the allocation, if, given μ_0 , one of the following sets of conditions is met:

- $\mu_a = 0, T_a > 0, EU_{r,a} = c, EU_{n,a} < c$
- $\mu_a = 1, T_a > 0, EU_{r,a} < c, EU_{n,a} = c$
- $\mu_a \in [0, 1], T_a > 0, EU_{r,a} = c, EU_{n,a} = c$

This definition of equilibrium is very standard and is identical to the notion of equilibrium I used in section 2.2.3. I can now write down the expected utility that each agent gains from entering the IPO allocation, given the level of rationing that occurs.

$$EU_{r,a} = \begin{cases} \frac{\Phi_\chi S}{(1-\mu_a)T_a} \chi + \frac{(1-\Phi_\chi)S}{T_a} \chi & \text{if } S < T_a(1-\mu_a)\delta \\ \Phi_\chi \delta \chi + \frac{(1-\Phi_\chi)S}{T_a} \chi & \text{if } T_a(1-\mu_a)\delta < S < T_a \delta \\ \delta \chi & \text{if } T_a \delta < S \end{cases}$$

$$EU_{n,a} = \begin{cases} \frac{(1-\Phi_\chi)S}{T_a} (\chi + E[\rho_a | \rho_a \geq -\chi]) & \text{if } S < T_a(1 - \mu_a)\delta \\ \frac{(1-\Phi_\chi)S}{T_a} (\chi + E[\rho_a | \rho_a \geq -\chi]) & \text{if } T_a(1 - \mu_a)\delta < S < T_a\delta \\ (1 - \Phi_\chi)(\chi + E[\rho_a | \rho_a \geq -\chi]) & \text{if } T_a\delta < S \end{cases}$$

In the above equations, $S < T_a(1 - \mu_a)\delta$ is the condition that requires rationing of shares for all values of ρ_a : enough rational participants have entered that they alone require the IPO to be rationed. $T_a(1 - \mu_a)\delta < S < T_a\delta$ implies that there are not enough rational traders to force rationing for all values of ρ_a , but when ρ_a is high and the noise traders bid the total demand will exceed supply. Similarly, $T_a\delta < S$ means that there aren't enough potential investors to induce rationing for any value of ρ_a .

I can now solve for the allocation participation equilibrium under definition 5. In doing so I will assume that one of the entry equilibria is played in the aftermarket, so that both agents receive the same level of expected utility after the allocation is over.

Lemma 8 *The Fixed-Price Allocation Entry Equilibrium*³¹: *If $\chi \geq \chi^{min}$, then total participation is described by*

$$\begin{aligned} T_a &= \frac{S}{c_a} [(1 - \Phi_\chi)(\chi + \bar{\rho}) + \sigma_\rho^2 \phi_\chi] \\ \frac{\partial T_a}{\partial \chi} &= \frac{S}{c_a} (1 - \Phi_\chi) > 0 \end{aligned}$$

while composition is described by

$$\begin{aligned} \mu_a &= 1 - 1_{\chi > \chi^*} \left[\frac{\chi \Phi_\chi}{(1 - \Phi_\chi) \bar{\rho} + \sigma_\rho^2 \phi_\chi} \right] \\ \frac{\partial \mu_a}{\partial \chi} &= 1_{\chi > \chi^*} \left[\frac{-\Phi_\chi + \mu_a \chi \phi_\chi}{(1 - \Phi_\chi) \bar{\rho} + \sigma_\rho^2 \phi_\chi} \right] \text{ except at } \chi = \chi^* \\ \frac{\partial \mu_a}{\partial \chi} &< 0 \quad \text{if } \mu_a > \frac{\Phi_\chi}{\chi \phi_\chi} \text{ and } \chi > \chi^* \end{aligned}$$

³¹The values of χ^{min} and χ^* are uninformative and can be found in the appendix.

If $\chi < \chi^{min}$, then there is no entry by either agent. Moreover, $\chi^{min} < \chi^$.*

As a corollary, if there is entry by the rational agent, then there is rationing for all values of ρ_a and entry by the noise agent. If there is entry by the noise agent, then there is rationing for all values of ρ_a greater than $-\chi$.

Because the noise trader learns his private value ρ_a after entering but before placing a bid, he has an option on that signal. This option has a positive value, but it also means that the noise trader does not always bid on objectively underpriced stocks. Thus both the rational and noise traders have situations where they can make excess profits: when the signal, ρ_a is high, the noise trader can bid and reap benefits unavailable to other participants, but when ρ_a is low, he does not bid allowing the rational trader to fully acquire an underpriced asset. The equilibrium allocation composition is a result of both traders paying the same entry costs – the excess profits of the noise and rational traders must be equal to the cost of participation.

The gain achieved by the rational trader is derived from underpricing, while the gain achieved by the noise trader is a composite of his own signal and of underpricing. By changing the price at which investors can purchase the stock, the firm can change the value of the different agents' gains. By increasing underpricing, the firm increases the gains of the rational trader relatively more than those of the noise trader, inducing more rational trader entry. At some level of underpricing, however, it becomes the case that the noise trader will almost always exercise his option. If this is true, then he too will realize almost the entire value from underpricing and this will drive the rational trader out of the allocation.

Figures 2-3 and 2-4 show how underpricing will affect allocation participation. The kink in the curve results from the fact that rational traders will not participate at all unless there is some non-zero level of underpricing. How much this is depends on their cost of entry. With a near zero cost, rational traders will participate to at least some degree with near zero underpricing. Noise traders, on the other hand, may actually participate in allocations with overpriced shares because of the option they hold on their private value.

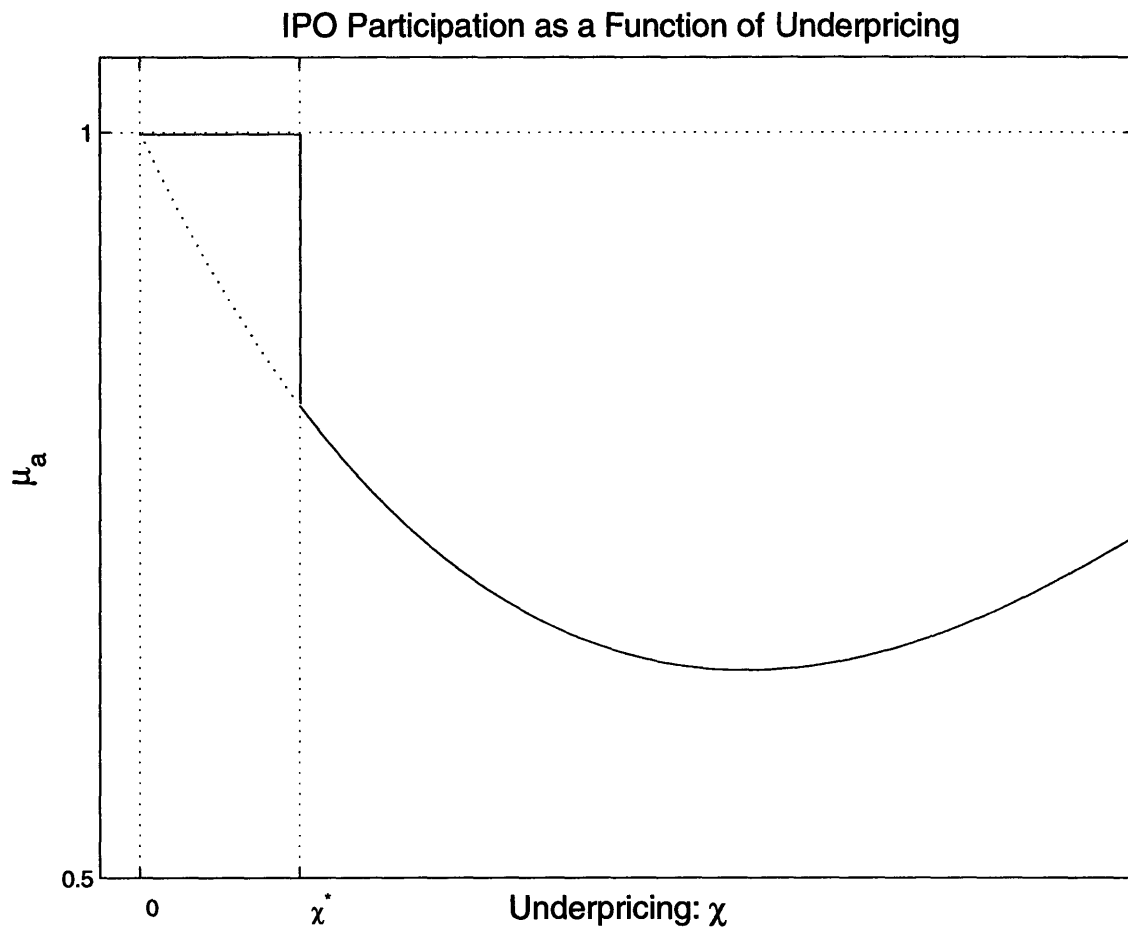


Figure 2-3: The value of μ_a obtained by underpricing. As the investment bank or firm reduces the asking price in the IPO allocation below the price that will exist in the aftermarket, the fraction of rational participants increases, up to a point.

Correspondence between Participation and Underpricing in the Allocation

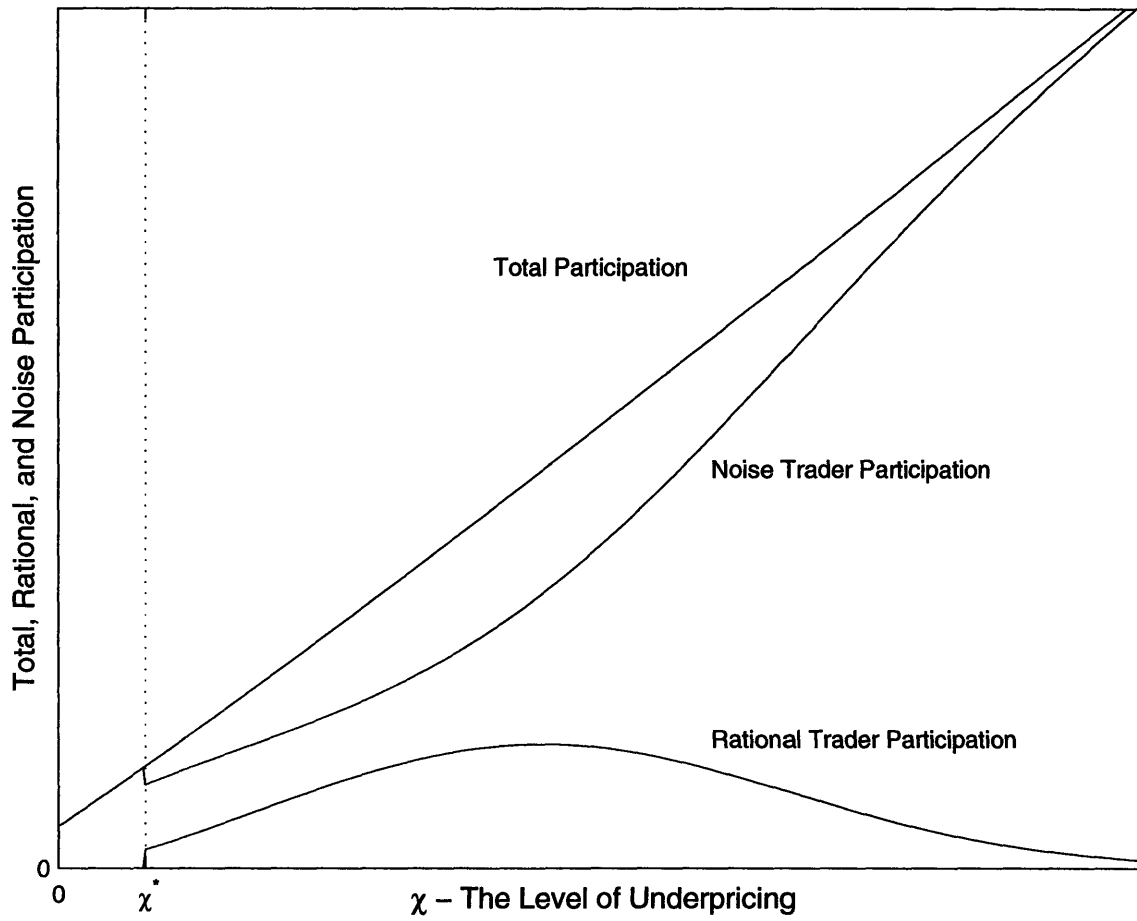


Figure 2-4: Participation as a result of underpricing.

The last detail of the allocation equilibrium is that the allocation is never partially subscribed for any value of ρ_a . To understand this, conduct a thought experiment where p_a and χ are held constant and c is increased. When c is low, agents are willing to enter because the value they get from a part of the allocation (δ shares) is equal to the loss they incur from the positive probability of receiving nothing. As c increases, they require a higher probability of being awarded the shares, and so fewer agents enter and the rationing rules are relaxed. Eventually, however, the rationing rules are fully relaxed so that even for high values of ρ_a , when the noise trader bids, each agent receives δ shares of stock. At this point, the agents must receive δ shares with probability 1 to offset the cost of entry. If the cost of entry is further increased, there is no way to reward investors enough that they are willing to participate – they cannot receive more than δ shares, or receive them with a probability greater than 1. As a result, if the allocation is beneficial enough that the rational trader will participate, it is also beneficial enough that his bidding will allow for a full allocation of shares when ρ_a is low. Similarly, if the allocation is beneficial enough that the noise trader participates, it is also beneficial enough that his bidding will allow for a full allocation of shares when ρ_a is high. As a result, for any given value of ρ_a , the allocation is never partially subscribed – it is either fully subscribed or there are no participants.

It is the case here, as in the aftermarket, that changing S does not change investor composition, only total participation. The mechanism is much the same as in the aftermarket: setting participation sets gain per investor. Since the latter is determined in equilibrium, changes in total supply will simply be offset by changes in total participation not by changes in composition.

2.3.3 The Firm's Problem

The firm's goal is to choose p_a so as to maximize proceeds from the allocation:

$$p_a = \operatorname{argmax} E p_0 - \pi \tag{2.10}$$

where π was defined in section 2.3.2 to be the expected value of underpricing, with $\pi = \chi + \frac{\alpha\delta}{2}Vp_0$. Because there are two aftermarket equilibria, Ep_0 and Vp_0 take on two values, and so they are step functions of p_a . To solve this problem, the most important step is to find the maximum value of p_a that will create the desirable aftermarket equilibria.

Because I want to demonstrate how the firm can create a high quality aftermarket, I will impose an aftermarket equilibrium selection criteria: the aftermarket equilibria played by investors is always that which minimizes the value to the firm. The aftermarket can be in either the μ^H or μ^L equilibria, but because the μ^H equilibrium has a lower average stock price, I assume that this is the resulting equilibrium whenever it is feasible. The firm can only force the μ^H equilibrium by making it the only feasible outcome. This equilibrium selection can be thought of as nature's move, if nature were malicious, but the point is to demonstrate that by affecting composition in the allocation, the firm can affect composition in the aftermarket. The firm is not manipulating beliefs in a clever way; while that would be part of any optimal mechanism, I want to show that the firm can affect the aftermarket without relying on specific institutional details. So, to make my point strongly, I will bias the aftermarket against the firm.

The firm has an open auction alternative to allocating shares through a mechanism with pricing discretion. To avoid complicated modelling with little intuition, I will assume that this amounts to taking the IPO directly to the aftermarket. Because the aftermarket is a rational expectations equilibrium where investors submit demand curves, it is an open auction. The only difference between it and any other auction is that investors can submit negative demands. While allowing short sales lowers the price conditional on entry, allowing traders to exploit one another more fully will increase entry which will increase the price. This also removes any underpricing in my open auction alternative, making it a desirable baseline. To simplify the discussion, I will assume that investors can go short so that an open auction replicates the aftermarket.

The open auction alternative will net the firm proceeds per share equal to the value of the price in the aftermarket when $\mu = \mu^H$. As a result, in order for the price discretionary procedure to be a viable choice, it must be the case that it allows the firm to set $p_a \geq Ep(\mu^H)$ in equilibrium while still maintaining investor participation.

To see how such an equilibrium can be constructed, let us assume that the firm sets $p_a = Ep(\mu^H)$, so that there is exactly zero underpricing with respect to the bad aftermarket equilibrium, but a positive amount of underpricing with respect to the good equilibrium. If rational investors believe that the bad equilibrium will prevail, then no rational investors will participate because they require some positive level of underpricing to justify spending the c_a cost of entry. The firm, then, is faced with the choice of withdrawing its IPO or continuing with the allocation. If it withdraws from the IPO, it receives zero proceeds and a zero payoff. If it continues with the IPO, and the noise traders receive a high signal, it may be able to place all of its shares with the noise traders. If, on the other hand, the noise traders receive a low signal, they will not bid and the firm will be unable to place its shares. Because the firm pays an infinite penalty if it cannot place all of its shares, and the noise traders will bid with positive probability, the firm cannot choose to continue the allocation. As a result, the firm will never choose to continue the allocation if rational investors believe that the bad aftermarket equilibrium will be played. Because the aftermarket only exists if the IPO goes forward, this means that the bad aftermarket equilibrium can never be played.

It only remains to be seen if the good aftermarket equilibrium, μ^L is feasible if the firm prices at $p_a = Ep(\mu^H)$. In this case,

$$\chi = Ep(\mu^L) - a\frac{\delta}{2}Vp(\mu^L) - Ep(\mu^H)$$

Because

$$\lim_{\chi \rightarrow \infty} T_a = \infty \text{ and } \frac{\partial T_a}{\partial \chi} > 0$$

these must exist a minimum χ such that either $\mu_a T_a > \mu^H T^H$ or $(1 - \mu_a) T_a >$

$(1 - \mu^H) T^H$.³² Remember that in the aftermarket, the equilibrium with a higher composition of noise traders, μ^H , has fewer investors of both types. Let this minimum value of underpricing be labelled χ^{**} . If $\chi > \chi^{**}$, then by construction, the equilibrium characterized by μ^H is not feasible, and so the aftermarket must play μ^L . Therefore, if

$$\chi^{**} \leq \text{Ep}(\mu^L) - a \frac{\delta}{2} \text{Vp}(\mu^L) - \text{Ep}(\mu^H)$$

then by naming $p_a = \text{Ep}(\mu^H)$, makes the μ^H aftermarket infeasible, and so the only remaining possibility is the μ^L aftermarket. As a result, the firm has created a high quality (low μ) aftermarket, and it receives at least as much in proceeds as what it would have received in the open auction. Moreover, if

$$\chi^{**} < \text{Ep}(\mu^L) - a \frac{\delta}{2} \text{Vp}(\mu^L) - \text{Ep}(\mu^H)$$

then the firm can receive strictly greater proceeds from a discretionary allocation than from an open auction.

To the contrary, if

$$\chi^{**} > \text{Ep}(\mu^L) - a \frac{\delta}{2} \text{Vp}(\mu^L) - \text{Ep}(\mu^H)$$

then the firm cannot create a high quality aftermarket without setting the price lower than would it would obtain in an open auction.

Alternately, one can look at the allocation of rents across groups to determine whether the firm will want to use an open auction or price discretionary allocation. The total rents received by investors is $c_a T_a$, which includes their risk aversion. As a result, if $c_a T_a < S (\text{Ep}(\mu^L) - \text{Ep}(\mu^H))$, meaning that investors capture less than the

³²Whether the firm creates the desirable aftermarket equilibrium on the noise trader margin or rational trader margin makes little difference to the basic intuition. If the firm has brought in more rational traders than the undesirable equilibrium can support, then more noise traders will enter in the aftermarket, but the good equilibrium will be preserved. Conversely, if the firm has brought in more noise traders than the bad equilibrium can support, more rational traders will enter in the aftermarket. The firm is only managing one aspect of participation – the rest is done by investors themselves when they make their entry decisions.

total gain from obtaining a high quality aftermarket, then the firm will prefer the fixed price mechanism to the open auction.

So what remains is to determine under what conditions the firm will prefer a discretionary allocation to an open auction. Many parameters have an ambiguous effect. The two that do not are c and c_a . A higher cost of entry in the aftermarket increases $\mu^H - \mu^L$ and so makes picking the right equilibrium more valuable to the firm. Decreasing c_a makes entry into the allocation more attractive to investors, and so the firm will have to underprice to a lesser degree to attract enough investors to create the good aftermarket.

I have shown that using an allocation procedure with price discretion can allow a firm to create a high quality secondary market in its stock. The result is based on the fact that by underpricing, the firm manipulates the payoffs of the various investors, inducing changes in their participation decisions. In fact, a firm can raise additional revenues by underselling its own aftermarket if doing so allows it to create a better aftermarket. By underpricing, the firm distributes rents to its investors, inducing them to take action that create the rents the firm is distributing. In this way underpricing is simply the cost of obtaining a high quality aftermarket and overcoming the coordination problem in forming a public market.

2.3.4 Discussion

My modelling is also consistent with some stylized facts about “flipping” – the rapid turnover of shares – and about institutional allocation. In the aftermarket, noise traders will always hold more of the stock on a per investor basis than rational traders because of their optimism. However, in the allocation, because shares are rationed equally, the individuals of different investor types will hold, on average, the same amount. The result is that in the first day of the aftermarket, rational investors will sell to noise traders. Since the variation in the first trading day’s price derives from the realization of private values, I also predict that rational traders will sell more to the noise traders when the first day’s return is higher. Underpricing and the

relationship between underpricing and rational sales mean that the rational traders are selling more into larger price increases – they are providing liquidity by “flipping”. While in any particular case the firm would like to prevent the rational traders from selling, the provision of liquidity is valuable in equilibrium. This partly explains the attitude of investment banks to flipping: those banks repeatedly allocate more shares to institutions, and institutions flip their shares more than other groups. Moreover, institutions flip more in “hot” IPOs than in “cold” ones, and “explicit penalty bids are rarely assessed against flippers” (Aggarwal (2003) and Aggarwal et. al (2002)).

2.4 Extension to an OLG Aftermarket

I will now extend the two period model of sections 2.2.1 through 2.2.3 into an OLG setting. Investors still live for two periods, but when old they are forced to liquidate their holdings. Aggregate stock supply in any given period is still S , but now the older generation is subject to price risk because of the noise trader risk that the stock bears. This noise trader/OLG setting is very similar to that of De Long et. al (1990), but the participation decision is entirely different.

The purpose of this extension is to show how noise traders can persist in the aftermarket and to show how the firm responds to a continuum of aftermarket choices. Because a stock’s price will contain the discounted value of all noise trader risk, the persistence of market composition will increase the value of the initial IPO placement. Moreover, because there is a continuum of aftermarket choices, I can analyze the firm’s marginal value to underpricing.

To clarify my notion of equilibrium, I will define the path of the economy and what is necessary for the a path to be an equilibrium:

Definition 6 *Path*: *A path is a sequence, $\{\mu^t, T^t\}$, that describes participation in the market for all time periods. I will show that T_t can be determined from the sequence of μ_t , so I will occasionally refer to a path simply as a sequence of $\mu_t, \{\mu^t\}$.*

Equilibrium Path: A sequence $\{\mu^t, T^t\}$ is an equilibrium path if each pair (μ_t, T_t) is an entry equilibrium.

This definition just says that each point on an equilibrium path is in equilibrium with respect to entry given every other point on the path.

2.4.1 The OLG Aftermarket

Now, instead of using $V_{d,t+1}$ to represent the risk of holding the stock, I will link composition across periods by replacing $V_{d,t+1}$ with $\mu^2 \sigma_\rho^2 \frac{1}{R^2} + \sigma_d^2$. Using this substitution, lemma 4 becomes

Lemma 9 (μ_t, T_t) is an entry equilibrium if:

The Corner Solution: If $c > C_1$, then $\mu_t = 1$ and $T_t = \frac{aS}{\sqrt{2c}} (\sigma_d^2 + \frac{1}{R^2} \sigma_\rho^2)$.

The Interior Solution: If $C_{min} < c$, then

$$\mu_t = \frac{1}{2} + \frac{1}{2} \frac{\bar{\rho}}{\sigma_\rho} \sqrt{\frac{8c \left(\frac{1}{R^2} \mu_{t+1}^2 \sigma_\rho^2 + \sigma_d^2 \right)}{\sigma_\rho^2 + \bar{\rho}^2} - 1} \quad (2.11a)$$

$$T_t = T_t(\mu_t, \mu_{t+1}) = \frac{2aS}{R^2} \frac{\mu_{t+1}^2 \sigma_\rho^2 + \sigma_d^2 R^2}{\sigma_\rho^2 + \bar{\rho}^2} \frac{\bar{\rho}}{2\mu_t - 1} \quad (2.11b)$$

if $\mu_t, \mu_{t+1} \in (\frac{1}{2}, 1)$. Moreover,

$$\bar{\rho} > 0, T_t > 0 \Rightarrow \mu_t > \frac{1}{2} \quad (2.12)$$

Otherwise, there is no equilibrium³³.

where it should be understood that the various values of C are analogous to those mentioned in the two and three period aftermarket models. The exact values are in the appendix.

³³If $c < C_{min}$ there is no equilibrium because no amount of entry will reduce investors' utility gains to c .

Equation 2.12 means that with endogenous entry, those traders with a higher expected value to the stock must make up the majority of market participants: if noise traders are on average optimistic, there must be more of them.

Lemma 9 and equation 2.11 also imply that an equilibrium path can be described purely from the sequence of $\{\mu^t\}$. This is because total participation is determined by the split of gains from trade and by the total risk, both of which are functions of μ_t . Moreover, these paths are deterministic. In equilibrium, investors' beliefs about the strategies of future market participants must be correct. Because investors base their own entry decisions on these correct beliefs, total entry is a deterministic function of composition.

A steady state is a state of the economy that can be repeated, i.e. a $\{\mu^t\}$ with all μ_t equal. In this case, I will refer to a steady state by the value of μ that repeats in that state. Please note that for a steady state I only require that a path has all μ_t equal; I do not require that nearby paths eventually converge to the steady state. In fact, of the two steady states I will find, transition paths converge to only one of them.

To find steady state solutions, I first need to find fixed points for the formulas in equation 2.11. There are two such fixed points³⁴, which I label μ^H and μ^L , with $\mu^H \geq \mu^L$. These solutions are valid under the same cost assumptions as in lemma 9 and equation 2.12:

$$\begin{aligned} C_{min} < c < C_1 &\Rightarrow \frac{1}{2} < \mu^H \leq 1 \\ C_1 < c &\Rightarrow \mu^H = 1 \\ C_{min} < c < C_{\frac{1}{2}} &\Rightarrow \frac{1}{2} < \mu^L \leq 1 \end{aligned}$$

Figure 2-5 shows the range of $\mu^{H,L}$ and $T^{H,L}$ and how they change with the cost of entry.

³⁴The values of μ^H and μ^L are in the appendix for the interested reader.

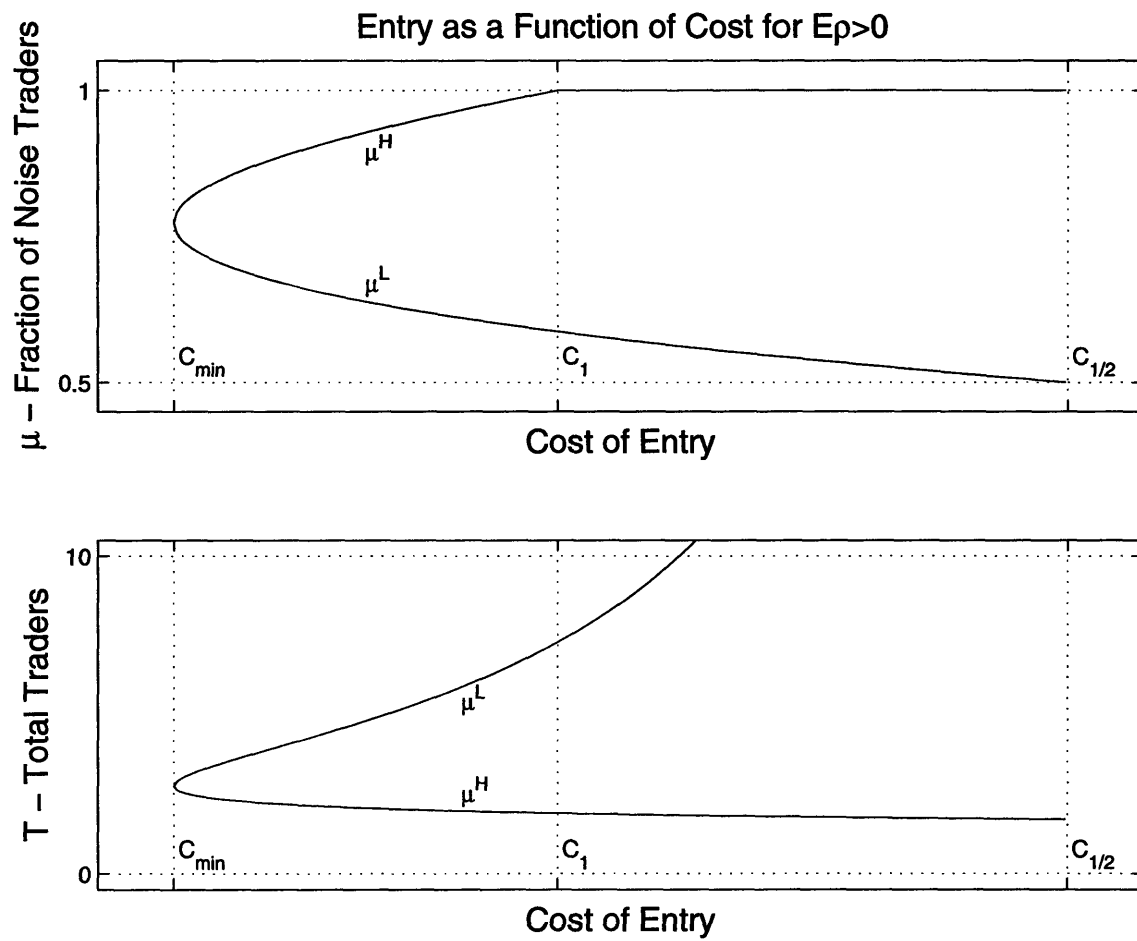


Figure 2-5: Participation decisions and market composition as a function of the cost of entry. The graphs are not calibrated, so the scale is not informative, except that $\frac{1}{2} < \mu < 1$.

The existence of steady states should no longer be surprising. I argued in section 2.2.3 that increasing the variance of the payoff to owning the stock encouraged proportionately more noise trader participation in the stock market. I argued that noise traders receive disproportionate gains to their expected utility from increasing the size of the risk premium, and so when that premium is higher, they will increase their relative participation. This all means that if there are more noise traders tomorrow, there will be more noise traders today because of the increase in noise trader risk and its compensation in the stock price.

In fact, μ^H and μ^L are analogous to the two equilibria in the three period model. The difference is that because I have extended the model to an OLG setting and allowed transition paths, what used to be equilibria in a setting without transition paths are now steady states.

Before examining transition paths, it remains to find which paths satisfy both the recursive relationships of lemma 9 and the definition of an equilibrium path.

Lemma 10 *Equilibrium Paths:* *If $\{\mu^t\}$ satisfies the recursive relationships in lemma 9, then it is an equilibrium path if and only if*

- $\frac{1}{2} < \mu^L \leq 1$ and $\mu_0 \in (\frac{1}{2}, \mu^H]$.
- $\mu^L < \frac{1}{2}$ and $\mu_t = \mu^H$.

These limitations on equilibrium paths are important because I have solved my economy recursively – there is nothing to pin down the value of μ_0 , except that it must belong to some equilibrium path. In fact, it will be the investment bank that picks the value of μ_0 so as to optimize IPO proceeds. Lemma 10 sets a boundary on which values of μ_0 that the investment bank can choose.

Clearly, now, I must know what happens along the various non-steady-state paths.

Lemma 11 *Convergence:* *If $\{\mu^t\}$ is an equilibrium path, and $\{\mu^t\} \neq \{\mu^H\}$, then*

$$\lim_{t \rightarrow \infty} \mu_t = \mu^L$$

Moreover, if $\{\mu^t\}$ and $\{\mu'^t\}$ are equilibrium paths, then

$$\mu_t > \mu'_t \Leftrightarrow \mu_{t+1} > \mu'_{t+1}$$

This lemma means that all transition paths converge to μ^L eventually and that paths do not cross. The non-crossing of paths means that picking any value of μ_t will pick the entire path, and when I discuss pricing it will mean that increasing the expected price at any given period increases it for all periods.

The existence of two steady states and the results of lemma 11 demonstrate persistence in the aftermarket. Because paths do not cross, more noise traders later means more noise traders now, so an observer of the marketplace would observe that stocks with high noise trader activity will see that activity level persist over time. In fact, this will make firms even more sensitive to their shareholder composition – because those compositions are persistent, they will have more impact on prices. The longer noise trader participation persists, the greater is the discounted value of risk.

Figure 2-6 illustrates the dynamics of μ_t over time for various starting values.

Pricing

Now that the evolution of the economy is understood, it remains to see how the one-period result that increased noise trader participation lowers expected prices is maintained. Lemma 9 and equation 2.3 together imply that

$$E p_t = \frac{1}{R} \left(\bar{d} + E \tilde{p}_{t+1} + \mu_t \bar{\rho} - (\bar{\rho}^2 + \sigma_\rho^2) \frac{\mu_t - \frac{1}{2}}{\bar{\rho}} \right) \quad (2.13)$$

Since we are interested in knowing how a stock's price varies with its ownership composition, we examine the derivative of $\frac{dE p_t}{d\mu_t}$ to see how the expected price changes with μ_t as one looks across equilibrium paths. Using the recursive form of μ_t from lemma 9,

$$\frac{dE p_t}{d\mu_t} = -\frac{\sigma_\rho^2}{R\bar{\rho}} \left(1 + \sum_{i=1}^{\infty} \frac{1}{R^i} \frac{d\mu_{t+i}}{d\mu_t} \right)$$

Evolution of the Current State over Time

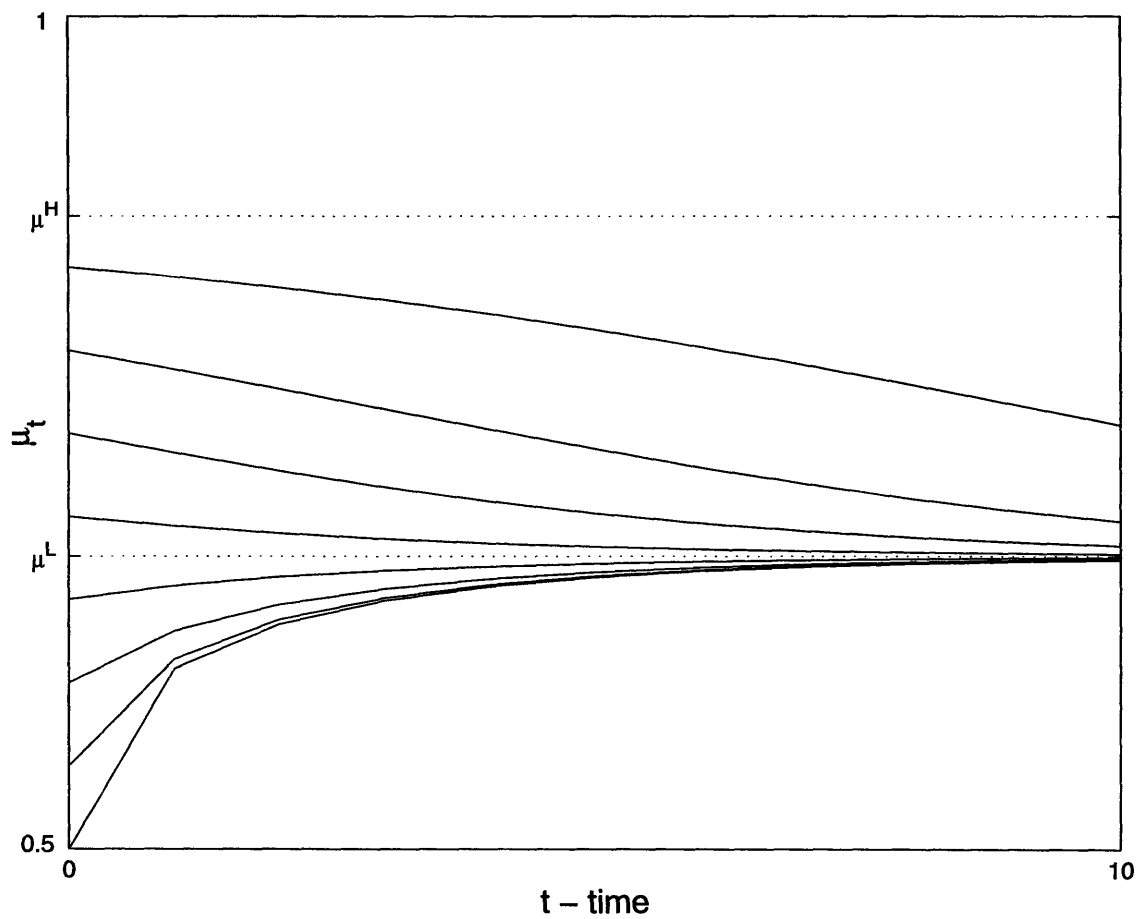


Figure 2-6: The evolution of μ_t over time from different starting values. Again, notice the two steady states, μ^H and μ^L . The paths that begin with $\mu_0 > \mu^H$ are not valid in equilibrium – they require some $\mu_t > 1$. There is nothing yet to tie down the value of μ_0 .

Since $\frac{d\mu_{t+i}}{d\mu_t} > 0$, the above expansion is always negative. So, path by path, the expected value of the price always decreases when the fraction of noise traders increases. This is true no matter what the optimism of the noise traders.

We can separate the effects of changing paths on the expected returns:

$$\frac{dE p_t}{d\mu_t} = \overbrace{-\frac{\sigma_\rho^2}{R\rho}}^{\text{Premium Effect}} \times \underbrace{\left(1 + \sum_{i=1}^{\infty} \frac{1}{R^i} \frac{d\mu_{t+i}}{d\mu_t}\right)}_{\text{Indirect Effect}}$$

The first effect is the “premium effect” and it is the same as in the one-period model. Total participation drops so quickly as noise traders enter that their indirect effect on the risk premium overwhelms their direct effect on demands.

The second effect is the “indirect effect”³⁵ is simply the result that changing market composition now changes the economy’s path, and so all future compositions are also changed. Because paths do not cross, increasing μ_t today also increases it tomorrow. Non-crossing paths also means that choosing μ_0 so as to increase current stock value will also act so as to increase future stock value. In this way, extending the model to multiple periods strengthens my pricing result.

Figure 2-7 shows numerically how the value of p_0 changes with μ_0 .

2.4.2 The Firm’s Problem

The firm’s goal is simply to maximize the expected proceeds from its IPO. It will do this by setting p_a so as to choose investor participation in the IPO, and as a result it will choose the equilibrium value of μ_0 seen in the aftermarket. By choosing a high quality aftermarket, it will be able to undersell the aftermarket price and still see higher proceeds than if it had chosen a low quality aftermarket and not undersold it. The key is that underpricing is relative to the aftermarket, not to some exogenously

³⁵The total evolutionary effect is always greater than 1, and may be greater than 2. In fact, there exists a $\hat{\mu} \in (\mu^L, \mu^H]$ such that if $\mu_t > \hat{\mu}$, then the evolutionary effect is greater than 2, and otherwise it is less than 2. In my model, this is just a reformulation of the convergence result of lemma 11.

Expected Opening Day Price as a Function of Market Composition

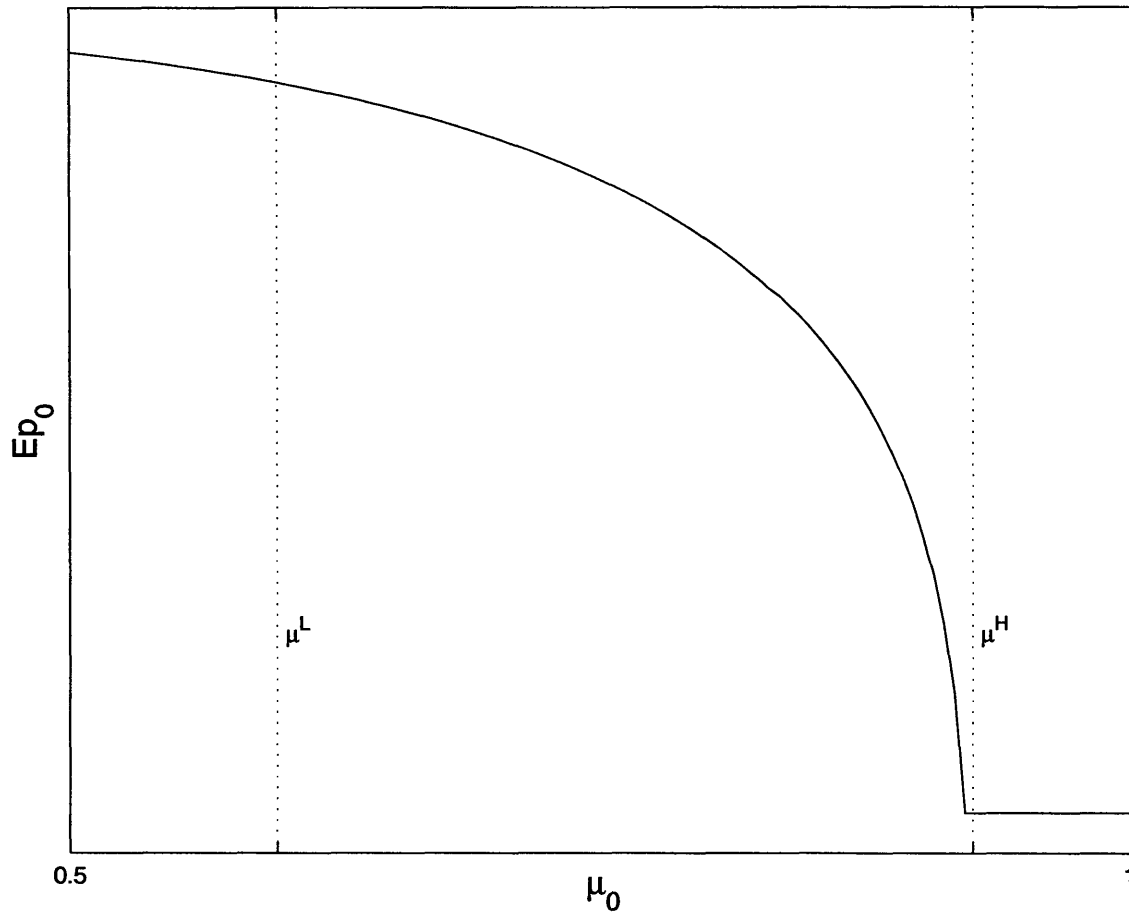


Figure 2-7: The change in the expected price as initial composition changes. There are no equilibrium paths when $\mu_0 > \mu^H$, so the price cannot be any lower than it is for μ^H .

given fundamental value. In conducting the allocation, the firm is making a sort of promise about what the aftermarket will look like. In creating a high quality aftermarket, the firm creates rents that it splits between itself and the investors. The rents to itself are excess proceeds while the rents to investors are returns on the first trading day: underpricing.

The investor's problem is the same as in section 2.3.2. That section only took account of investors' ability to forecast allocation and aftermarket participation. As a result, the investors' response to underpricing will be the same for any given aftermarket price.

In the infinite horizon aftermarket model, there are a continuum of equilibria, including those near $\mu_0 = \frac{1}{2}$ that also have T near infinity. As a result, the firm can never create enough entry to eliminate all equilibria but one without infinite underpricing. It can, however, set a boundary, where all feasible equilibria have $\mu_0 < \mu_0^*$. Doing this involves the same mechanism as in the three-period case – underprice the μ_0^* equilibrium to such an extent that the number of investors who participate makes any equilibria with $\mu_0 > \mu_0^*$ infeasible. How investors and the firm behave will then depend on their beliefs about what the aftermarket will look like. Instead of setting up an elaborate mechanism for enforcing beliefs, I will simply assume that all agents and the firm believe that $\mu_a = \mu_0$, so that the equilibrium played in the aftermarket will have the same composition as the initial allocation. I do this because I do not want to pretend that the fixed price allocation is the optimal mechanism for eliciting information about investor quality – as before, I simply want to show that the firm can manage its investors to its benefit.

The firm's goal is to maximize the proceeds from the allocation: to maximize p_a subject to the constraint that $\mu_a = \mu_0$. Allocation proceeds are

$$p_a = \mathbb{E}p_0(\mu_0) - \delta \frac{\sigma}{2} \mathbb{V}p_0(\mu_0) - \chi$$

So the firm's problem amounts to

$$p_a = \operatorname{argmax} \quad E p_0(\mu_0) - \frac{a\delta}{2} V p_0(\mu_0) - \chi$$

$$\text{s.t.} \quad \mu_a = \mu_0$$

Because the moments of the aftermarket price depend on μ_0 which is chosen by the firm and must equal μ_a , I can write

$$p_a = E p_0(\mu_a) - \delta \frac{a}{2} V p_0(\mu_a) - \chi$$

and have the firm choose χ so as to maximize p_a . In this problem, as opposed to the 3-period aftermarket examined in section 2.3, there are a continuum of possible aftermarkets. Taking derivatives, the first order condition is

$$1 = \left[-\delta a \mu_a \sigma_\rho^2 - \frac{\sigma_\rho^2}{R\bar{\rho}} \left(1 + \sum_{i=1}^{\infty} \frac{1}{R^i} \frac{d\mu_{t+i}}{d\mu_t} \right) \right] \frac{\partial \mu_a}{\partial \chi} \quad (2.14)$$

Before continuing, recall from lemma 8 that if $\hat{\chi}$ is such defined so that $\left. \frac{\partial \mu_a}{\partial \chi} \right|_{\hat{\chi}} = 0$, then the firm will never set a price such that $\chi > \hat{\chi}$. This because as χ increases beyond $\hat{\chi}$, the firm attracts a less favorable composition to its share allocation.

Because the firm must allocate all of its shares, it must always underprice by at least χ^* to induce rational traders to enter. Together with the global minimum of μ_a at $\hat{\chi}$, a firm using the fixed price allocation procedure will always set underpricing so that $\chi^* < \chi < \hat{\chi}$. Moreover, for $\chi^* < \chi < \hat{\chi}$, $\frac{\partial \mu_a}{\partial \chi}$ is positive and decreasing. As a result, as χ increases, the firm sees diminishing returns in composition to increasing underpricing.

The second part of the first order condition is $-\delta a \mu_a \sigma_\rho^2 - \frac{\sigma_\rho^2}{R\bar{\rho}} \left(1 + \sum_{i=1}^{\infty} \frac{1}{R^i} \frac{d\mu_{t+i}}{d\mu_t} \right)$, which represents the decline in pricing from increasing relative noise trader participation. The term is negative and decreasing in μ_a – so as underpricing increases and μ_a declines, the firm sees diminishing returns to changing investor composition. As a result, there is one or zero solutions to equation 2.14 within the range $\chi^* < \chi < \hat{\chi}$.

If there exists a solution to equation 2.14, then that is the desired level of χ in a fixed price allocation. If no such solution exists, then the firm will set $\chi = \chi^*$ according to its constraint.

It remains to ask which corporations will derive the most value from changing their ownership composition. This means looking at the sensitivity of the expected stock price to changes in σ_d . First, I assume that $\mu^L > \frac{1}{2}$ and look across equilibrium paths. Since $\frac{\partial}{\partial \sigma_d} \frac{d\mu_{t+i}}{d\mu_t} > 0$, then $\left| \frac{dE p_t}{d\mu_t} \right|$ must increase in σ_d .

So if a corporation faces an investor composition problem – $\mu^L < \frac{1}{2}$ means that there is only one aftermarket equilibrium – then when looking at a cross section of companies, it is those companies with high cash flow variance that stand to gain or loss the most from changes in their ownership structure. Low variance companies have prices that respond only a small amount to changes in μ – therefore, one would expect that companies would underprice in proportion to σ_d . This matches evidence from recent IPO studies, such as Ljungqvist and Wilhelm (2002), that indicate uncertain cash flows, particularly those in recent technology companies, will be associated with more underpricing.

2.4.3 Discussion

I have assumed throughout this section that a firm cares only about its issue price – the firm takes no account of any return trips to the capital markets. While this makes my results very stark, it is not entirely realistic. A firm may have several reasons why it would wish to create a high quality aftermarket even aside from proceeds raised during the IPO. For example, the firm may want to issue shares again later, or the firm's managers might hold a share of unissued stock. If the firm values the aftermarket for its own sake, it will underprice more than otherwise – that a firm can increase the quality of the aftermarket can be seen from figure 2-3. In fact, Schultz and Zaman (2001) find that during the technology bubble, corporate insiders sold fewer of their own shares and saw larger underpricing than in other industries. Moreover, Ljungqvist and Wilhelm (2002) find that in the same bubble the total size of post-

IPO insider ownership fell, but that programs that allow insiders to buy shares at the IPO price became much more common. These two papers provide evidence for my prediction that increasing a managers stake in an IPO would cause them to put more weight on the aftermarket and hence underprice more. This is a striking example of a time when it is not desirable to give a manager a large stake in a company.

As in the previous section, I have not provided conditions under which the firm will prefer an open auction to a procedure that uses price discretion. My goal here was to show how a firm will act when faced with a continuum of possible aftermarkets, and as a result, how its underpricing decision might change based on any firm characteristics. My two predictions – that if owners sell a small portion of their personal stakes or that if firms face very uncertain cash flows, they will underprice more – are two examples of the effects of firm characteristics.

The three period aftermarket demonstrates how a firm can use discretionary pricing to achieve a high quality aftermarket, but extending the model to a more dynamic setting allows me to show how and why a firm might choose between more varied aftermarket options. For example, I have shown that a firm can use underpricing to select an aftermarket, whereas now I am able to derive conditions on the optimal level of underpricing given different aftermarket choices. Moreover, it has allowed me to show that noise trader participation and risk can be persistent.

2.5 Conclusion

I have argued that there is a coordination problem in the formation of public trading markets. Because investors take as given the quality of the market when they intend to sell, they must premise their participation and purchasing decisions on the participation decisions of many other investors. In particular, noise traders will create the very risk that induces them to participate. As a result, a higher fraction of noise trader participants will be associated with a lower expected stock price. Furthermore, these participation decisions do not unravel when investors trade with one another –

noisy stock price encourage the entry of the very traders that cause the noise. Firm value is affected by market conditions, so the delegated allocation of IPO shares by investment banks dominates open auction formats because it allows the firm to control its initial ownership composition.

My analysis is relevant for many different types of markets and market formations. First, the flow of investors into or out of a given market means that market quality is constantly changing – in fact, the market is being constantly re-formed. The coordination problem thus exists in all markets to the extent that they allow entry. Second, while I have applied my analysis to the public offering of stocks, there is no reason in principle it could not apply to the public offering of other assets, such as sovereign debt, or Seasoned Equity Offerings.

2.6 Appendix

Parameter Values

From Sections 2.2.1, 2.2.2, and 2.2.3:

$$C_1 = \frac{(\sigma_\rho^2 + \bar{\rho}^2)^2}{8V_{d,t+1}\bar{\rho}^2}$$

$$C_{min} = \frac{\sigma_\rho^2 + \bar{\rho}^2}{8V_{d+1}}$$

Figure 2-1: $R = 1$, $\bar{\rho} = 1.4$, $\sigma_\rho = 1$, $a = 1$, $S = 1$, $\sigma_d = 1$, $c = .27$.

From Section 2.2.4:

$$C_{\frac{1}{2}} = \frac{(\sigma_\rho^2 + \bar{\rho}^2)(\sigma_\rho^2 + 8\sigma_d^2)}{8\sigma_d^2(\sigma_\rho^2 + 4\sigma_d^2)}$$

$$C_1 = \frac{(\sigma_\rho^2 + 2\sigma_d^2)^3(\sigma_\rho^2 + \bar{\rho}^2)^2}{32\sigma_d^2\bar{\rho}^2(\sigma_\rho^2 + \sigma_d^2)}$$

C_{min} is the minimum value across $\mu \in [\frac{1}{2}, 1]$ of $EU_{r,t}$ when evaluated at the T from lemma 6.

μ^H and μ^L are the solutions to the equation $EU_{r,t} = c$ when $EU_{r,t} = c$ is evaluated at the T from lemma 6.

C_{min} , μ^H , and μ^L all solve eighth order polynomials in μ .

From Section 2.3:

χ^{min} is defined by

$$c = \delta [(1 - \Phi_{\chi^{min}}) (\chi^{min} + \bar{\rho}) + \sigma_\rho^2 \phi_{\chi^{min}}]$$

while χ^* is defined by

$$c = \frac{\Phi_{\chi^*} \chi^*}{E[\rho_a | \rho_a \geq -\chi^*]} \delta (\chi^* + E[\rho_a | \rho_a \geq -\chi^*])$$

Figure 2-3: $\bar{\rho} = .5$, $\sigma_\rho = 10$.

Figure 2-4: $\bar{\rho} = .5$, $\sigma_\rho = 10$, $S = 1$, $c = 1$.

Section 2.4:

$$C_{min} = \frac{(\bar{\rho}^2 + \sigma_\rho^2) (\bar{\rho}^2 + \sigma_\rho^2) + 4R^2 \sigma_d^2}{16 \sigma_d^2 \bar{\rho}^2} - \frac{(\bar{\rho}^2 + \sigma_\rho^2) \sqrt{((\bar{\rho}^2 + \sigma_\rho^2) + 4R^2 \sigma_d^2)^2 - 16R^2 \sigma_d^2 \bar{\rho}^2}}{16 \sigma_d^2 \bar{\rho}^2}$$

$$C_{\frac{1}{2}} = \frac{1 (\bar{\rho}^2 + \sigma_\rho^2) R^2}{2 \sigma_\rho^2 + 4\sigma_d^2 R^2}$$

$$C_1 = \frac{1 (\bar{\rho}^2 + \sigma_\rho^2)^2 R^2}{8 (\sigma_\rho^2 + \sigma_d^2 R^2) \bar{\rho}^2}$$

with $C_{min} < C_{\frac{1}{2}}$ and $C_{min} < C_1$, and

$$\lim_{\sigma_d \rightarrow 0} C_{min} = \frac{1}{2} R^2$$

$$\mu^H = \frac{1 (\bar{\rho}^2 + \sigma_\rho^2) R^2}{2 (\bar{\rho}^2 + \sigma_\rho^2) R^2 - 2c\bar{\rho}^2} + \frac{R \bar{\rho} \sqrt{-(\bar{\rho}^2 + \sigma_\rho^2)^2 (R^2 - 2c) + 8c\sigma_d^2 ((\bar{\rho}^2 + \sigma_\rho^2) R^2 - 2c\bar{\rho}^2)}}{2 \sigma_\rho (\bar{\rho}^2 + \sigma_\rho^2) R^2 - 2c\bar{\rho}^2}$$

$$\mu^L = \frac{1 (\bar{\rho}^2 + \sigma_\rho^2) R^2}{2 (\bar{\rho}^2 + \sigma_\rho^2) R^2 - 2c\bar{\rho}^2} - \frac{R \bar{\rho} \sqrt{-(\bar{\rho}^2 + \sigma_\rho^2)^2 (R^2 - 2c) + 8c\sigma_d^2 ((\bar{\rho}^2 + \sigma_\rho^2) R^2 - 2c\bar{\rho}^2)}}{2 \sigma_\rho (\bar{\rho}^2 + \sigma_\rho^2) R^2 - 2c\bar{\rho}^2}$$

Figure 2-5: $R = 1, \bar{\rho} = 1.4, \sigma_\rho = 1, a = 1, S = 1, \sigma_d = 1.$

Figure 2-6: $R = 1, \bar{\rho} = 1.4, \sigma_\rho = 1, a = 1, S = 1, \sigma_d = 1, c = .27.$

Figure 2-7: $R = 1.05, \bar{\rho} = 1.4, \sigma_\rho = 1, a = 1, S = 1, \sigma_d = 1, c = .285.$

Outlines of Proofs

Below I outline the proofs of the lemmas contained in the paper. Full details are available upon request.

Lemmas 3 and 5: In the expressions for $EU_{i,t}$, substitute $T_t = n + r$ and $\mu_t = \frac{n}{n+r}$. For $i = n, r$, $\frac{\partial EU_{i,t}}{\partial i} < 0$.

Lemma 4: First, I look for corner solutions where only one investor chooses to enter: $\mu = 1$ and $\mu = 0$. If $\mu = 1$, then $EU_{n,t}|_{\mu_t=1} = T_t$ implies that $T_t = \frac{aS}{\sqrt{2c}}V_{d,t+1}$. Because $EU_{n,t}|_{\mu_t=1} > EU_{r,t}|_{\mu_t=1}$ if and only if $c > C_1$, the corner solution for $\mu_t = 1$ is as given in the lemma. Because $EU_{n,t}|_{\mu_t=0} > EU_{r,t}|_{\mu_t=0}$ for all values of c , there is no corner solution for $\mu_t = 0$.

Second, I look for interior solutions in which both traders choose to enter, and so $EU_{r,t} = EU_{n,t} = c$. This is a system of two equations and two variables with the solution given in the statement of the lemma. Because $\bar{\rho} > 0$ and $T_t > 0$, it must be the case that $\frac{1}{2} < \mu_t < 1$. Solving this inequality for c results in the cost condition $C_{min} < c < C_1$.

Lemmas 6 and 9 can be proved in an identical method to that for lemma 4.

Lemma 7: The first statement can be proved by substituting in the value of T from lemma 6 into equation 2.7 and noting that $\frac{\partial E_P(\mu)}{\partial \mu} < 0$.

The second statement can be shown by observing that $\frac{\partial \mu T(\mu)}{\partial \mu} < 0$ and $\frac{\partial (1-\mu)T(\mu)}{\partial \mu} < 0$.

Lemma 8: If $T_a \delta < S$, then $EU_{n,a} > EU_{r,a}$, so $\mu_a = 1$, and $EU_{n,a} = c$ implies $\chi = \chi^{min}$. So $\chi < \chi^{min}$ implies that there is no entry by either agent.

If $S < T_a \delta$, then $\chi > \chi^{min}$ and there is rationing for high values of ρ_a , so some noise traders must be participating. $EU_{n,a} = c$ implies T_a is as given in the statement of the lemma. If $\chi < \chi^*$, then $EU_{r,a} < c$ so there is no rational trader participation.

If $S < T_a(1 - \mu_a)\delta$, then $\chi > \chi^*$ and there is rationing for all values of ρ_a , so both traders are participating. $EU_{n,a} = c$ and $EU_{r,a} = c$ are two equations with two variables which can be solved for the values of T_a and μ_a given in the statement of the lemma.

Lemmas 10 and 11: If equation 9 is inverted, the result is $\mu_{t+1} = F(\mu_t)$, where F is increasing and convex. Therefore if $\mu_0 > \mu^H$, then the path will require for some t that $\mu_t > 1$, and so it is not an equilibrium path. If $\mu_0 < \mu^H$, then the path will converge to μ^L . But all values of μ_t must be greater than $\frac{1}{2}$ (lemma 9), so if $\mu_L < \frac{1}{2}$, $\mu_0 < \mu^H$ cannot be an equilibrium path.

The second statement in lemma 11 follows from the fact that F is monotonic.

Bibliography

- [1] Allen, Franklin; Gale, Douglas. 1994, Limited Market Participation and Volatility of Asset Prices, *The American Economic Review*, 84, 933-955.
- [2] Aggarwal, Reena; 2000, Stabilization Activities by Underwriters after Initial Public Offerings, *The Journal of Finance*, 55 (3), 1075–1103.
- [3] Aggarwal, Reena, 2003, Allocation of Initial Public Offerings and Flipping Activity, *Journal of Financial Economics*, 68, 111–135.
- [4] Aggarwal, Reena; Prabhala, Nagpurnanand R.; Puri, Manju. 2002, Institutional Allocation in Initial Public Offerings: Empirical Evidence, *The Journal of Finance*, 57, 1421–1442.
- [5] Biais, Bruno; Faugeron-Crouzet, Anne Marie. 2002, IPO Auctions: English, Dutch,...French, and Internet, *Journal of Financial Intermediation*, 11, 9–36.
- [6] Barberis, Nicholas; Shleifer, Andrei. 2003, Style Investing, *Journal of Financial Economics*, 68, 161–199.
- [7] Barberis, Nicholas; Shleifer, Andrei; Wurgler, Jeffery. 2003, Comovement, working paper.
- [8] Benveniste, Lawrence M.; Busaba, Walid Y. 1997, Bookbuilding vs. Fixed Price: An Analysis of Competing Strategies for Marketing IPOs, *The Journal of Financial and Quantitative Analysis*, 32, 383–403.

- [9] Benveniste, Lawrence M.; Busaba, Walid Y.; Wilhelm, William J. Jr.; 1996, Price Stabilization as a Bonding Mechanism in New Equity Issues, *Journal of Financial Economics*, 42, 223-255.
- [10] Benveniste, Lawrence M.; Spindt, Paul A. 1989, How Investment Bankers Determine the Offer Price and Allocation of New Issues, *Journal of Financial Economics*, 24, 343–362.
- [11] Benveniste, Lawrence M.; Wilhelm, William J. 1990, A Comparative Analysis of IPO Proceeds Under Alternate Regulatory Environments, *Journal of Financial Economics*, 28, 173–207,
- [12] Bergstresser, Daniel; Poterba, James. 2003, Asset Allocation and Asset Location: Household Evidence from The Survey of Consumer Finances, working paper.
- [13] Bertaut, Carol C. 1998, Stockholding Behavior of U.S. Households: Evidence from 1983-1989 Survey of Consumer Finances, *Review of Economics and Statistics*, 80, 263275.
- [14] Booth, James R.; Chua, Lena. 1996, Ownership Dispersion, Costly Information, and IPO Underpricing, *Journal of Financial Economics*, 41, 291–310.
- [15] Cespa, Giovanni. 2002, Short-Term Investment and Equilibrium Multiplicity, *European Economic Review*, 46, 1645–1670.
- [16] Chemmanur, Thomas J.; Liu, Huanliang (Mark). 2002, How Should a Firm Go Public? A Dynamic Model of the Choice Between Fixed-Price Offerings and Auctions in IPOs and Privatizations, working paper.
- [17] De Long, J. Bradford; Shleifer, Andrei; Summers, Lawrence H.; Waldman, Robert J. 1990, Noise Trader Risk in Financial Markets, *The Journal of Political Economy*, 98 (4), 703–38.

- [18] Derrien, Francois; Womack, Kent L. 2003, Auctions vs. Bookbuilding and the Control of Underpricing in Hot IPO Markets, *The Review of Financial Studies*, 16, 31–61.
- [19] Ellis, Katrina; Michaely, Roni; O'Hara, Maureen. 2000, When the Underwriter Is the Market Maker: An Examination of Trading in the IPO Aftermarket, *The Journal of Finance*, 55 (3), 1039–1074.
- [20] Ellul, Andrew; Pagano, Marco. 2003, IPO Underpricing and After-Market Liquidity, working paper.
- [21] Goetzmann, William N.; Kumar, Alok. 2003, Diversification Decisions of Individual Investors and Asset Prices, working paper.
- [22] Goyal, Amit; Santa-Clara, Pedro. 2003, Idiosyncratic Risk Matters!, *Journal of Finance* 58, 9751007.
- [23] Grinblatt, Mark; Keloharju, Matti. 2001, How Distance, Language, and Culture Influence Stockholdings and Trades, *Journal of Finance*, 56, 10531073.
- [24] Jeanne, Olivier; Rose, Andrew K. 2002, Noise Trading and Exchange Rate Regimes, *The Quarterly Journal of Economics*, 117, 537–569.
- [25] Jenkinson, Tim; Ljungqvist, Alexander. 2001, *Going Public: The Theory and Evidence on How Companies Raise Equity Finance*, Second Edition, Oxford University Press, New York.
- [26] Hanley, Kathleen Weiss. 1993, The Underpricing of Initial Public Offerings and the Partial Adjustment Phenomenon, *Journal of Financial Economics*, 34, 231–250.
- [27] Kelly, Morgan. 1995, All Their Eggs in One Basket: Portfolio Diversification of US Households, *Journal of Economic Behavior and Organization*, 27, 8796.

- [28] Ljungqvist, Alexander P.; Nanda, Vikram; Singh, Rajdeep. 2003, Hot Markets, Investor Sentiment, and IPO Pricing, working paper.
- [29] Ljungqvist, Alexander P.; Wilhelm, William J. Jr. 2002, IPO Pricing in the Dot-com Bubble, *The Journal of Finance*, 58, 723–752.
- [30] Loughran, Tim; Ritter, Jay R. 2002, Why Don't Issuers Get Upset About Leaving Money on the Table in IPOs?, *The Review of Financial Studies*, 15, 413–443.
- [31] Lowry, Michelle; Schwert, G. William. 2002, IPO Market Cycles: Bubbles or Sequential Learning?, *The Journal of Finance*, 57, 1171–1200.
- [32] Odean, Terrance. 1998, Are Investors Reluctant to Realize Their Losses?, *Journal of Finance* 53, 1775–1798.
- [33] Pagano, Marco. 1989, Endogenous Market Thinness and Stock Price Volatility, *The Review of Economics Studies*, 56, 269–287.
- [34] 2002, Ritter, Jay; Welch, Ivo. 2002, A Review of IPO Activity, Pricing, and Allocations, *The Journal of Finance*, 57, 1795–1828.
- [35] Rock, Kevin. 1986, Why New Issues Are Underpriced, *Journal of Financial Economics*, 15, 187–212.
- [36] Sherman, Ann E. 2002, Global Trends in IPO Methods: Book Building vs. Auctions, working paper.
- [37] Sherman, Ann; Titman, Sheridan. 2002, Building the IPO Order Book: Underpricing and Participation Limits with Costly Information, *Journal of Financial Economics*, 65, 3–29.
- [38] Schultz, Paul; Zaman, Mir. 2001, Do the Individuals Closest to Internet Firms Believe They are Overvalued?, *Journal of Financial Economics*, 59, 347–381.

- [39] Spatt, Chester; Srivastava, Sanjay. 1991, Pre-play Communication, Participation Restrictions, and Efficiency in Initial Public Offerings, *The Review of Financial Studies*, 4, 709–726.
- [40] Spiegel, Matthew. 1998, Stock Price Volatility in a Multiple Security Overlapping Generations Model, *The Review of Economic Studies*, 11, 419–447.
- [41] Stoughton, Neal M.; Zechner, Josef. 1998, IPO-Mechanisms, Monitoring, and Ownership Structure, *Journal of Financial Economics*, 49, 45-77.
- [42] Welch, Ivo. 1992, Sequential Sales, Learning, and Cascades, *Journal of Finance*, 47, 695–732.

Chapter 3

Heterogeneous Beliefs and the Principal-Agent Problem

Written with Tobias Adrian.

3.1 Introduction

While academics have examined the principal-agent problem in great detail, there has been little work on contracting between agents with heterogeneous priors and on how this affects labor market conditions and outcomes. In this paper, we argue that a principal will select an agent with beliefs different from his own, and that the side-bets in which they engage form an important part of the contract's value. The principal faces a trade-off because any side-bets distort the agent's optimal choice of effort, but these distortions can be optimally chosen. In fact, the ability to engage in side-bets with the agent makes the principal ambivalent about learning: experience allows the agent to better manage a project and so increase its profitability, but it also reduces the distance between his beliefs and those of the principal. Depending on labor market and industry conditions, the principal may optimally choose to rapidly hire and fire his agents rather than let their experience accumulate.

Such an approach is necessary because reconciling the functioning of some business

activities, for example an asset trading floor, with standard labor economics is very difficult. One of the most difficult tasks is to understand what makes a good and valuable trader. While traders are unsurprisingly unwilling to reveal their thought processes, they do share certain rules of behavior. The most common rule, repeated frequently by managing directors, is that they must “take a view” – that they should take very aggressive positions rather than more moderate positions more likely to turn a moderate profit. From anecdotal experience, it seems preferable that they fail extravagantly rather than succeed moderately. On the trading floor, no attempt is made to force traders to form common expectations about financial markets or the economy. A newly hired trader is not educated in order to align the traders views with the views of other agents or the principle. Another striking fact is that job tenure at a specific firm is very short. New traders are hired frequently, others are fired, and often the recently fired are hired at a different institution for a similar but slightly different job.

A second question is why bonuses based on firm performance, such as stock options, are common at low levels in a firm’s hierarchy. With homogeneous beliefs, it seems unlikely that this is an incentive device, as lower level employees have small individual contributions to firm profitability. At the same time, those employees can purchase options in a public market so it is hard to argue that they provide a special opportunity to exploit behavioral biases. In a world with heterogeneous beliefs, remunerating employees with stock options is a cheap way of providing incentives.

This paper derives the optimal contract as well as the optimal hiring and firing decision by principals of agents with heterogeneous priors about the average payoff of an investment project. Hiring agents with a different belief is valuable for the principle as he can trade state payoffs with the agent – he can “sell dreams”. However, over time, the heterogeneity in priors is likely to disappear due to learning, while at the same time the agent is likely to become more productive due to increased experience. In the hiring decision, the principle faces a trade-off between keeping an agent to benefit from increased productivity and hiring a new agent in order to

exploit differences in beliefs.

Our main theoretical contribution of the paper is to show that linear contracts are optimal when priors are heterogeneous for risk neutral agents. In standard models, when the agent and the principle are risk neutral, selling the firm to the agent is an optimal contract, but this is not the case in a model with differing priors. Instead, the incentive schedule depends on the degree of disagreement between the principle and the agent. The principle exploits the disagreement by offering overconfident agents a contract with higher-powered incentives; even though this provides more than first-best effort, the principal is compensated through his expected payoffs from the side-bets. Because optimistic agents reflect an opportunity to the principal, he always gains from their existence – they could always just not engage in side bets. This is not true, however, when the agent is pessimistic because in this case the principal has to pay more to overcome the low expectations of the agent.

In regards to the examples above, “taking a view” can simply be interpreted as a managing director’s directive to traders to be less risk averse and more extreme in their thinking, both of which enable side-bets. The turnover is a natural consequence to this, as the model describes. The stock options held by low level employees are interpreted as a form of side-bet: whether or not the optimistic agent believes his project constitutes a large part of firm profitability, and whether or not he can buy options on the side, he always thinks that non-vested options are more valuable than the less optimistic principal believes them to be.

The trade-offs embodied in our model are applicable to a number of incentive problems in the financial sector beyond the trading floor. For example, hedge fund investors can choose among a variety of hedge funds with different views of the world; the reallocation of capital corresponds to the hiring/firing decision in our model. Venture capitalists invest in entrepreneurs with biased beliefs, and have to choose how fast to reallocate their funds. The model can also be used to study how the optimal incentive schedule of CEOs should depend on their priors, and how frequently CEOs should be fired. The model is also applicable to the junior academic market:

assistant professors are churned frequently, and only a small fraction of them are granted tenure.

In addition, we make a number of new cross-sectional and time series predictions regarding the functioning of labor markets. The first is that proxies for heterogeneous beliefs, high-powered incentives, and labor market turnover should all be correlated. Second, labor market churning and high powered incentives should be associated with high industry profits. Both of these can be broken down in more detail, which we do in section 3.4.

The principle-agent problem originates in the work of Wilson (1969), Spence and Zeckhauser (1971) and Ross (1973). The source of the problem is that the principle cannot contract the effort of the agent directly, but only the outcomes. In general settings, the optimal contract that the principle offers to the agent is nonlinear. Holmstrom and Milgrom (1987), bothered by the fact that most contracts in the real world are fairly simple, formulate a dynamic, continuous time principle-agent model with CARA utility and show that the optimal contract is linear in such a setting. We adopt the general framework of Holmstrom and Milgrom, and introduce heterogeneous beliefs as well as labor market churning.

There is a large theoretical literature studying the impact of heterogeneous beliefs on asset prices. Miller (1977) and Harrison and Kreps (1978) show that the combination of heterogeneous beliefs and short selling constraints can lead to a speculative component in asset prices, a point that has recently been studied more extensively by Scheinkman and Xiong (2004). Consistent with these theoretical papers, Diether, Malloy and Scherbina (2002) and Scherbina (2004) show that the dispersion of analyst forecasts predict excess stock returns. Gilchrist, Himmelberg and Huberman (2003) show that the mispricing induced by heterogeneous beliefs and short selling constraints affects the investment behavior (in physical assets) of firms. Harris and Raviv (1993) argue that the sheer size of trading volume is indirect evidence for the importance of heterogeneous beliefs.

In the current paper, we take the existence of heterogenous beliefs as exoge-

nous, but there are papers that address the choice of beliefs explicitly. Benabou and Tirole (2002) model the psychological trade-off between self-serving beliefs and self-deception, deriving the optimal degree of belief distortions. Brunnermeier and Parker (2004) study the trade-off between overconfidence and distorted choices in a dynamic setting. Gervais and Odean (2001) study both the emergence and survival of overconfidence of asset traders.

A number of recent studies have shown that biased beliefs change the optimal contract in a principal-agent setting. Gervais, Heaton and Odean (2003) and Kleiber (2003) point out that it is cheaper for a principal to provide incentives when the agent is overconfident. Bolton, Scheinkman and Xiong (2003) study the optimal contract when asset prices are distorted due to overconfidence of investors.

There are a few papers analyzing the impact of heterogeneous beliefs on corporate financing decisions empirically. Landier and Thesmar (2003) use survey data that include questions about expectations, and can thus construct measures of overconfidence. They demonstrate that optimistic entrepreneurial beliefs lead to a preference for short-term debt, excessive risk-taking, and an increased likelihood of default. Bertrand and Schoar (2003) investigate whether and how individual managers affect corporate behavior and performance. They track managers across different firms over time and find that manager fixed effects matter for a wide range of corporate decisions. Malmendier and Tate (2003) find that investment is significantly more responsive to cash flow if the CEO displays overconfidence. What these studies have in common is to show that heterogeneous beliefs matter for firms decision making.

A number of studies empirically investigate the impact of heterogeneous beliefs on trading behavior. Osler and Oberlechner (2003) survey currency traders on major fixed income trading floors and find overconfidence in two forms: traders tend to overestimate professional success compared to their bosses' estimate, and underestimate forecast uncertainty of exchange rates. They find that the degree of overconfidence does not have an impact on trading profitability, but it is correlated with seniority on the floor. Mann and Locke (2001) examine the trading activity of professional futures

traders. They find that more successful traders are more likely to be overconfident and take on more risky positions. These studies matter for our theory as they justify our main assumption, namely, that agents have different beliefs than their principles.

The remainder of the paper is organized as follows. In sections 3.2 and 3.3, the model is introduced and analyzed, and the main theorems are discussed. In section 3.4, we derive and discuss empirical predictions. Section 3.5 concludes.

3.2 The Model

Since our objective is to analyze the economic mechanisms associated with contracting between agents with different beliefs, we use a simple model for parsimony.

We place our principal agent model in continuous time in order to derive the optimal contract; in fact, we find that the optimal contract in our initial setup is linear in the relevant variables. If the reader wishes to avoid the formalism of continuous time, he or she can consult appendix 3.6.1 for a discrete time version of the problem. If we constrain the contract in the discrete time version to be linear, then the solution to that problem is the same as the solution to the continuous time problem. All of our extended results and analysis follow from the form of the contract rather than from the continuous time formalism.

3.2.1 Assumptions

Information Structure

We consider a continuous-time, finite-horizon economy. The uncertainty of the economy is described by a one-dimensional, standard Brownian motion B_t for $0 \leq t \leq 1$, defined on a complete probability space (Ω, F, P) , where F is the augmented filtration generated by B_t .

Opportunities

There is a single risky project in the economy which pays a dividend D_1 at time 1. D_1 is the value of the process D_t at time 1, where $D_0 = 0$. The rights to the project are owned by the principal, but to undertake the project he hires an agent. The agent exerts an unobservable level of effort, e_t . In choosing e_t , the agent changes the evolution of the project's payoff to

$$dD_t = (\mu + e_t) dt + \sigma dB_t \quad (3.1)$$

In return for the agent's labor, the principal offers the agent a payment, S . The principal cannot observe e_t , but he can observe the path of D_t .

The project has two parameters: μ and σ , which measure the project's initial mean and its standard deviation. Both μ and σ are constant and are assumed to be positive.

Strategies

The agent's choice of effort must satisfy standard integrability conditions to avoid pathologies. Effort, e_t , must be a real number and must be F_t measurable, meaning it can depend on the entire history of the economy up to time t .

The contract, S , must be of the form

$$S = S_1 + \int_0^1 a(t, D_t) dt + \int_0^1 b(t, D_t) dD_t \quad (3.2)$$

where both a and b must satisfy standard integrability conditions. Moreover, both $a(t, D_t)$ and $b(t, D_t)$ must be F_t measurable.

Beliefs and Preferences

Both the principal and agent are risk neutral over consumption.

We assume that the principal and agent use different probability measures, so their contract will reflect their differing expectations about the rate of change of D_t .

The principal has probability measure P , and he believes that D_t evolves according to equation 3.1. The principal receives expected utility equal to

$$EU_p = E_0^P [D_1 - S] \quad (3.3)$$

We have apparently given the principal objectively correct beliefs, but this is only because we wish to use the principal as our point of reference. Through section 3.2 we make no assumptions about what set of beliefs is correct, while in section 3.3 we will assume that the principal has learned enough over time so that P is the objective probability measure.

The agent, on the other hand, has probability measure Q under which

$$dB_t = \frac{\delta}{\sigma} dt + dB_t^Q$$

and believes that D_t evolves according to

$$dD_t^Q = (\mu + \delta + e_t) dt + \sigma dB_t^Q \quad (3.4)$$

We have used δ to parameterize the degree to which the beliefs of the principal and of the agent differ. If $\delta > 0$, we say the agent is optimistic, and this should be taken to mean that he is optimistic relative to the principal. Because δ is assumed to be constant, the probability measure of the agent Q is absolutely continuous with respect to that of the principal P , i.e., both traders agree on zero-probability events. Let $\xi_t \equiv (dQ/dP)_t$ denote the Radon-Nikodym derivative of the probability measure Q with respect to P . Then

$$\xi_t = e^{-\frac{1}{2} \frac{\delta^2}{\sigma^2} t + \frac{\delta}{\sigma} B_t}$$

The agent pays a quadratic cost for the effort

$$c(t, D_t, e_t) = \frac{1}{2} c e_t^2$$

and so receives expected utility from the contract equal to

$$EU_a = E_0^Q \left[S - \int_0^1 \frac{1}{2} ce_t^2 dt \right] = E_0^P \left[\xi_1 S - \xi_1 \int_0^1 \frac{1}{2} ce_t^2 dt \right] \quad (3.5)$$

We further assume that the agent has an outside option of \bar{u} .

It is worth noting that we assume that the principal and agent do not update their beliefs as a result either of coming into contact with one another or of gathering additional data during the duration of the contract. We do this to avoid running afoul of the no-trade-theorem of Milgrom and Stokey (1982). In particular, if the principle has private information, a contract that the agent accepts might not exist.

Our principal and agent stick to their differing beliefs throughout the course of the contract, and in this sense they are irrational. It should be understood, however, that we take no stand on how the difference in beliefs arise, only that they do not change during the life of the contract. Later, in section 3.3, we allow the participants to update their beliefs at discrete intervals.

3.2.2 Equilibrium

The equilibrium in this economy can be solved analytically, and the optimal values of S and e can be computed in closed form. We will use the first order approach, which will yield a unique solution.¹ The interested reader who wishes to read an in-depth discussion of why the first order approach works in continuous time set-ups should see appendix A in Schattler and Sung (1993).

Equilibrium: An equilibrium consists two functions: a function $e^*(t, D)$ that is F_t measurable and integrable, and a function S^* of the form

$$S^* = S_1^* + \int_0^1 a^*(t, D_t) dt + \int_0^1 b^*(t, D_t) dD_t$$

¹Grossman and Hart (1983), Hart and Holmstrom (1987), and Schattler and Sung (1993) discuss the validity of the first order approach. It is valid in our framework because both the principal's and agent's problems are strictly convex.

where $a^*(t, D_t)$ and $b^*(t, D_t)$ are F_t measurable and integrable. They must satisfy

$$\{e_t^*\} \in \operatorname{argmax} E_0^Q \left[S^* - \int_0^1 \frac{1}{2} c e_t^{*2} dt \right] \quad (3.6)$$

$$E_0^Q \left[S^* - \int_0^1 \frac{1}{2} c e_t^{*2} dt \right] \geq \bar{u} \quad (3.7)$$

$$\{S^*\} \in \operatorname{argmax} E_0^P \left[D_1 + (\xi_1 - 1) S - \xi_1 \int_0^1 \frac{1}{2} c e_t^*(S)^2 dt \right] \quad (3.8)$$

$$dD_t = (\mu + e_t^*) dt + \sigma dB_t$$

which are, respectively, the agents incentive compatibility constraint, the agent's participation constraint, the principal's optimality condition, and the dividend process.

In order to solve both the principal's problem and the agent's problem we will make use of the following lemma:

Lemma 12 *Under standard integrability conditions, if*

$$V_t = \max_{\{e_t\}} E_t \left[\int_t^1 z(s, D_s, e_s) ds \right]$$

exists, and

$$dD_t = (\mu + e_t) dt + \sigma dB_t$$

then V_t is such that

$$V_t = V_0 - \int_0^t z(s, D_s, e_s) ds + \int_0^t \nabla V_s \sigma dB_s$$

and

$$e_t^* = \operatorname{argmax}_{e_t} z(t, D_t, e_t) + e_t \nabla V_s$$

This just means that integrals of that kind can be maximized point-wise, with a correction for the diffusion of the value function. The proof is in appendix 3.6.2.

The Agent's Problem

Because the agent has outside option \bar{u} , he will accept the principal's contract S only if the expected gain (equation 3.5) exceeds \bar{u} . Assuming he accepts the contract, the agent's problem is to find $\{e_t^*\}$ such that

$$\{e_t^*\} \in \operatorname{argmax} \mathbb{E}^Q \left[S - \int_0^1 \frac{1}{2} c e_t^2 dt \right] \quad (3.9)$$

subject to the constraint

$$dD_t^Q = (\mu + \delta + e_t) dt + \sigma dB_t^Q$$

To solve this, notice that

$$\mathbb{E}^Q \left[S - \int_0^1 \frac{1}{2} c e_t^2 dt \right] = \mathbb{E}^Q \left[\int_0^1 \left(a(t, D_t) + b(t, D_t)(\mu + \delta + e_t) - \frac{1}{2} c e_t^2 \right) dt \right]$$

where we have used the fact that $\mathbb{E}^Q \left[\int_0^1 b(t, D_t) \sigma dB_t^Q \right] = 0$. Using lemma 12, we see that

$$b(t, D_t) - c e_t^* + \nabla V_t = 0 \quad (3.10)$$

and that

$$\begin{aligned} dV_t^Q &= - \left(a(t, D_t) + b(t, D_t)(\mu + \delta + e_t) - \frac{1}{2} c e_t^2 \right) dt + \nabla V_t \sigma dB_t^Q \\ &= -dS^Q + \sigma c e_t^* dB_t^Q + \frac{1}{2} c e_t^{*2} dt \end{aligned}$$

which can be integrated to show

$$S_1^Q = \bar{u} + \int_0^1 \frac{1}{2} c e_t^{*2} dt + \int_0^1 \sigma c e_t^* dB_t^Q \quad (3.11)$$

since both S_0^Q and V_1^Q are equal to zero. The agent's participation constraint will be tight, implying that $V_0^Q = \bar{u}$.

We can now match terms, using equations 3.11 and 3.2, to show that

$$e_t^* = \frac{1}{c}b(t, D_t) \quad (3.12)$$

Together, equations 3.12 and 3.11 give us the agent's optimal effort choice given the salary form in equation 3.2. They also give us the value that the salary has to have in order to implement a particular level of effort. Because the agent's problem is strictly convex and results in a unique optimal control, the salary given in equation 3.11 will implement the effort choice given in equation 3.12. This allows us to use the first order approach, substituting the agent's decisions into the principal's problem.

The Principal's Problem

The principal must offer the agent a contract S that the agent will accept, and it must maximize the expected proceeds to the principal. The principal's problem is:

$$\max_{\{S\}} E_0^P [D_1 - S] \quad (3.13)$$

$$\text{s.t. (i) } dD_t = (\mu + e_t) dt + \sigma dB_t$$

$$\text{(ii) } \{e_t^*\} \text{ solves the agent's problem}$$

$$\text{(iii) } E^Q \left[S - \int_0^1 \frac{1}{2} ce_t^2 dt \right] \Big|_{\{e_t^*\}} \geq \bar{u}$$

To begin, notice that

$$S_t = \bar{u} + \int_0^t \sigma ce_s^* dB_s^Q + \int_0^t \frac{1}{2} ce_s^{*2} ds = \bar{u} + \int_0^t \sigma ce_s^* dB_s^P + \int_0^t \frac{1}{2} ce_s^{*2} ds - \int_0^t c\delta e_s^* ds \quad (3.14)$$

The last term, $\int_0^t c\delta e_s^* ds$ shows that there is a complementarity between the principal's choice of the effort level to implement (e_t^*) and the agent's beliefs (δ). The extra value

that the agent places on the contract equals

$$\int_0^t c\delta e_s^* ds = \int_0^t b(t, D_t)\delta ds$$

In other words, because the agent believes the dividend will grow more quickly, he over-values that portion of his contract that rewards him for changes in the dividend. While this does not induce him to spend more effort (equation 3.12 does not depend on any particular measure), it does induce him to put a higher value on contracts that are steeper in D_t , while at the same time he exerts a higher effort as a result of those same contracts.

Our model differs from the standard principal-agent model in two important respects, both originating in the difference between the principal's and agent's probability measures. The first is that the principal will assess the agent's payoff differently than does the agent. This is important because the standard solution to the risk-neutral problem is to "sell the firm" to the agent: $S = D_1 - \alpha$ for some constant α . That solution works because if the principal and agent have the same probability measures, then the agent will pick the level of e_t that maximizes total welfare. Different probability measures means that the agent no longer behaves this way.

The extra value that the agent places on the contract can be interpreted as a kind of side-bet. Because the principal and agent disagree on how likely a given state is, they both gain from trading on their beliefs. If, for example, the agent is optimistic, the principal would like to offer the agent a contract that gives the agent a higher payment in particularly good states; the agent will overvalue those states, relative to the principal, and so the principal will be able to offer slightly lower payments in other states that he considers more likely.

The key point, however, is that the principal must use one contract both to elicit the optimal amount of effort and to create the most desirable set of side-bets. Selling the agent high states has a cost in that the agent will pick a higher level of effort; increasing effort incurs a direct, quadratic cost, as per equation 3.9. Because this cost is quadratic, the principal will eventually reach a point where it is no longer profitable

to engage in further side-bets and we arrive at equilibrium.

To solve the principal's problem, notice that

$$\mathbb{E}[D_1 - S] = \mu - \bar{u} + \mathbb{E} \left[\int_0^1 e_t^* dt - \int_0^1 \frac{1}{2} c e_t^{*2} dt + \int_0^1 c \delta e_t^* dt \right]$$

where we have used the fact that $\mathbb{E} \left[\int_0^1 (\sigma c e_s^* + \sigma) dB_s^P \right] = 0$. Using lemma 12, this is maximized at

$$e_t^* = \frac{1}{c} (1 + c\delta) + \frac{1}{c} \nabla V_t$$

where ∇V_t now refers to the principal's value function. Note that because ∇V_t is independent of measure, and hence independent of effort choice, we can substitute e_t^* back into the principal's problem to get

$$\mathbb{E}[D_1 - S] = \mu - \bar{u} + \mathbb{E} \left[\int_0^1 (1 + c\delta + \nabla V_t) \frac{1}{2c} (1 + c\delta - \nabla V_t) dt \right]$$

which is maximized, state by state, at $\nabla V_t = 0$. As a result, the principal's total expected utility is

$$\mu + \frac{1}{c} (1 + \delta c)^2 - \bar{u} \tag{3.15}$$

and the optimal contract sets

$$b^*(t, D_t) = 1 + c\delta \tag{3.16}$$

To find the value of $a(t, D_t)$, we can proceed by matching terms from equations 3.2 and 3.14 under P . We find that

$$a^*(t, D_t) = \frac{1}{2} c e_t^{*2} - c \delta e_t^* - (\mu + e_t^*) b(t, D_t) = -(\mu + \delta) (1 + c\delta) - \frac{1}{2c} (1 + c\delta)^2 \tag{3.17}$$

The result is an equilibrium in which we can see the tradeoff between side-bets and incentive compatibility without any of the complications induced by risk sharing. Because the agent's participation constraint binds exactly, maximizing total welfare

is equivalent to maximizing the principal's utility.

Proposition 8 *The solution to the principal's problem is a linear contract of the form $S = A + BD_1$ where*

$$A^* = \bar{u} - (\mu + \delta)(1 + c\delta) - \frac{1}{2c}(1 + c\delta)^2 \quad (3.18a)$$

$$B^* = 1 + c\delta \quad (3.18b)$$

This contract is pareto optimal and the principal receives

$$EU_p = \mu + \frac{1}{2c}(1 + c\delta)^2 - \bar{u}$$

The principal's problem is equivalent to one in which he attempts to maximize social welfare. The sum of utilities is

$$\begin{aligned} EU_p + EU_a &= E \left[D_1 + (\xi_1 - 1)S - \int_0^1 \frac{1}{2} ce_t^2 dt \right] \\ &= E \left[\int_0^1 \left(\mu + e_t + \delta b(t, D_t) - \frac{1}{2} ce_t^2 \right) dt \right] \end{aligned} \quad (3.19)$$

The principal faces a trade-off between rents received from side-bets and rents received from effort on the part of the agent. Given a linear contract $A + BD_1$, total welfare equals

$$\mu + e^* + B\delta - \frac{1}{2} ce^{*2} \quad (3.20)$$

The total value of the side-bets is included in the term $B\delta$, and it can be increased by raising B . As implied by equation 3.12, however, this comes at a cost:

$$-\frac{1}{2} ce^{*2} = -\frac{1}{2} \frac{B^2}{c} \quad (3.21)$$

which increases rapidly in B . So second-best contracting forces the principal to choose between side-bets and effort in a way that prevents him from fully exploiting the differences in measure between himself and the agent.

The contract is linear because the ability of the principal to engage in side-bets is proportional to slope of the contract. The value to the side bets is

$$E_0^P [(\xi_1 - 1) S] = E_0^P \left[\int_0^1 \delta b(t, D_t) dt \right]$$

The principal faces a tradeoff between linear gains to side-bets and effort, and a quadratic cost to effort; the resulting contract is linear.

Because the principal will exploit his ability to engage in side-bets with the agent until the marginal benefit from doing so equals the marginal cost to the project, the principal actually gains less utility from the project itself than he would without the side-bets. To see this, we first must assess the value of the agent's compensation under the principal's measure; in other words, how much value is the principal obtaining through side-bets with the agent?

$$E_0^P U_a = E^P \left[A + B^* D_1 - \frac{1}{2} c e^{*2} \right] = \bar{u} - \delta (1 + \delta c) \quad (3.22)$$

Here, $\delta (1 + \delta c)$ represents the value to the principal of the side-bets with which he engages the agent.

We next compute the value to the principal of engaging an agent for whom $\delta = 0$. Following proposition 8, this value is

$$EU_p|_{\delta=0} = \mu + \frac{1}{2c} - \bar{u}$$

So, after re-arranging, equation 3.15 becomes

$$EU_p = \left[\mu + \frac{1}{2c} - \bar{u} \right] + [\delta (1 + \delta c)] + \left[-\frac{\delta^2 c}{2} \right] \quad (3.23)$$

and is composed of three terms. The first is the value to the principal of contracting with an agent with whom he shares beliefs. The second term represents the gains to the principal from the fact that he can engage in side-bets with the agent – it represents the gross benefit the principal receives from the difference in measure

between himself and the agent. The third is as yet accounted for, and it represents the cost to the principal, in terms of project profitability, of engaging in side-bets. These side-bets distort the contract and the associated effort choice away from what would otherwise be optimal, and this distortion is costly.

A Comparison to the First-Best

There is no first best solution to the principal's problem. To see this, let us examine how the total welfare of equation 3.19 evolves:

$$\left(\mu + e_t + \delta b(t, D_t) - \frac{1}{2} c e_t^2 \right) dt$$

There is a clear optimal value for e_t , but the principal wishes to engage in an unbounded number of side-bets with the agent: increasing $b(t, D_t)$ has no cost. Because total surplus equals

$$E_0^P \left[D_1 + (\xi_1 - 1) S - \xi_1 \int_0^1 \frac{1}{2} c e_t^2 dt \right]$$

the principal would like to set $S = \infty$ when $\xi_1 > 1$ and $S = -\infty$ when $\xi_1 < 1$. This is a function of risk neutrality – the principal and agent have no risk sharing considerations, so are quite willing to trade an infinite number of claims in state where they value differently. It is only the fact that the side-bets come at the cost of suboptimal effort in the second best case that allows for equilibrium.

3.2.3 Discussion

In standard principal-agent models with a risk-neutral principal and agent, the optimal contract involves the principal “selling the project” to the agent. The optimal contract is $D_1 - \alpha$. This solution accomplishes the first best because without risk sharing considerations, the principal and agent agree on what effort level is optimal, and the agent will choose that level. When the principal and agent are allowed to place side-bets with one another, this solution is no longer valid. In fact, with

heterogeneous beliefs, there is no unconstrained first-best solution.

There is a second-best solution, and it involves an optimal trade-off between side-bets and incentives for effort. In the risk neutral case, however, it involves more than selling the firm to the agent: $B = 1 + c\delta > 1$. This is not a contract we commonly see people signing, and the reason is that no one is truly risk neutral over their labor income. It does occasionally happen, however, that agents sign contracts where they effectively a larger marginal bonus than the produce in marginal profit. This can be the case when, for example, salespeople are compensated based on revenues, or traders are compensated on trading profits without accounting for costs. In fact, the compensation for traders rarely includes provisions for support and other costs.

Lastly, cost of effort, c , acts to drive contracts away from $B = 1$: $\frac{\partial B^*}{\partial c} = \delta$, $B(\delta = 0) = 1$. Recall from equation 3.21 that the marginal welfare cost to sharpening the incentive schedule is inversely proportional to the cost of effort: the more expensive the marginal unit of effort, the less the elastic is the agent's choice. The source of loss to the principal from engaging in side-bets is the distorted choice of effort by the agent, but this distortion is less when effort is very costly. As a result, the principal is free to offer a sharper incentive schedule the greater is c .

3.3 The Labor Market

To analyze how heterogeneous beliefs might affect hiring and firing decisions, we will now embed the problem of section 3.2 in a simplified labor market. To maintain simplicity, we will assume that there is a single monopolist principal and many available agents. In a market with competition the gains from trade due to the heterogeneity in beliefs would be somehow allocated between the principal and agent. The particular way that gains were split would determine many of the market details, but we wish to focus on broad aspects of the marketplace instead.

3.3.1 Assumptions

We will assume

- At each time $t \in T = \{0, 1, 2, \dots\}$, N potential agents are in the marketplace with priors parameterized by δ_t^i for $i = \{1, 2, \dots, N\}$. We assume the priors are ordered such that $\delta_t^1 \geq \delta_t^2 \geq \dots \geq \delta_t^N \geq -\frac{1}{c}$. These δ_t^i apply equally to all projects.
- The pool of agents is renewed each period so that $\{\delta_t^1, \delta_t^2, \dots, \delta_t^N\}$ are always available, plus any agents that have previously been hired and have returned to the labor market pool.
- All agents have an outside option valued at \bar{u} .
- At times $t \in T$, the project matures, payments are made, and $D_{t \in T}$ is reset to zero for the next period.
- There is one principal and he can hire only one agent per period.
- The market game each period proceeds as follows: The principal offers a menu of one-period contracts. This menu is unconstrained, save that each contract may only be contingent on the final payoff of the project. Then each agent matches with a contract if he so desires. The principal then chooses which contract/agent pair he wishes to engage.

We will further assume that an agent learns over time in two ways if he is employed. First, his beliefs converge to those of the principal. In particular, if an agent is first employed at time t' with beliefs $\delta_{t'}^i$ and forecast error $\gamma_{t'}^i$, then at every $t > t' \in T$, he updates his beliefs according to the Kalman-Bucy filter²

$$\delta_t^i = \frac{\delta_{t'}^i(\sigma^2/\gamma_{t'}^i) + \sigma(B_t - B_{t'})}{(\sigma^2/\gamma_{t'}^i) + (t - t')} \quad (3.24)$$

²This filter is derived in appendix 3.6.3.

For simplicity, we assume that all potential agents have the same prior $\gamma_t^i = \gamma_0$ $\forall t$ before being hired (or re-hired). It would be straightforward (but algebraically cumbersome) to introduce heterogeneity in γ^i across agents.

Since the principal is permanent, we assume for the purposes of the labor market that he has learned everything he will learn – the principal knows μ with certainty. This is stronger than what we assumed in section 3.2; there we were only concerned with differences in beliefs, not with learning. However, this is not a major shortcoming, as the following results only depend on the relative distance of the agents and the principal's beliefs. What matters for our analysis is that their beliefs converge over time, but they do not necessarily have to converge to the true value.

Second, we assume that agents acquire a form of principal-specific human capital. If an agent is hired at time t' and the base profitability (before effort) of a project he undertakes is μ , then, if the agent has been continuously employed until time $t \in T$, his base profitability is assumed to be $\mu + \eta(t - t')$. We assume that the agent loses this accumulated increase in profitability if he is fired.

The last two assumptions reflect two types of learning that the agent undergoes. The first is a function of memory and revising beliefs: as the agent gains experience, he learns more and more about the profitability of the project he can undertake (he learns about the true probability distribution). The second assumption is a function of skills: as the agent gains familiarity with the project, he becomes better at undertaking it and he raises the base profitability of the project. Henceforth, we will refer to the two types of learning as *memory* and *skills*.

3.3.2 Equilibrium

By examining the principal's expected utility from a one period contract of an agent i that has been employed for t periods, equation 3.15,

$$\mu + \eta t + \frac{1}{2c} (1 + c\delta_t^i)^2 - \bar{u}$$

we can see that the principal benefits from the skills of his employees, but he also benefits from heterogeneous beliefs. As a result, the principal would ideally like his agents to gain familiarity with the project so as to increase the average payoff (η), but without learning more about the profitability of the project and lowering δ . Unfortunately, this isn't possible.

The principal will have one of two dominant strategies, depending on the outstanding pool of potential agents. The first potential dominant strategy is to hire the most optimistic available agent every period, and to fire him at the end of the period. This takes advantage of the principal's ability to engage in side-bets, and provides him with a constant stream of willing participants. The second potentially dominant strategy is to hire the most optimistic available agent and to continue re-hiring him forever. This takes advantage of the agent's ability to learn about the project and increase its profitability.

To assess the tradeoffs, we calculate the payoffs from the two potentially dominant strategies. If he hires and fires an agent of type i every period, then the principal will obtain

$$\frac{1}{1-\beta} \left[\mu + \frac{1}{2c} (1 + c\delta_0^i)^2 - \bar{u} \right] \quad (3.25)$$

If the principal hires an agent i forever, his payoff is

$$\sum_{t=0}^{\infty} \beta^t \left[\mu + t\eta + \frac{1}{2c} E_0^P (1 + c\delta_t^i)^2 - \bar{u} \right] \quad (3.26)$$

where we take account of agent i 's learning process. If the agent follows the optimum Kalman-Bucy filter, and the forecast variance of his prior is γ_0 , then

$$\delta_t^i = \frac{\delta_0^i \sigma^2 / \gamma_0 + \sigma B_t}{\sigma^2 / \gamma_0 + t} \quad (3.27)$$

so that

$$E_0^P (1 + c\delta_t^i)^2 = \left(1 + c\delta_0 \frac{\sigma^2 / \gamma_0}{\sigma^2 / \gamma_0 + t} \right)^2 + \frac{c^2 \sigma^2 t}{(\sigma^2 / \gamma_0 + t)^2} \quad (3.28)$$

If

$$N(\delta_t^i) = (1 - \beta) \sum_{t=0}^{\infty} \beta^t \mathbb{E} (1 + c\delta_t^i)^2$$

and if we define $EUC(\delta_t^i)$ as the gain from churning over that of hiring forever, then

$$EUC(\delta_t^i) = \frac{1}{1 - \beta} \left[\frac{1}{2c} (1 + c\delta_t^i)^2 - \frac{\beta}{1 - \beta} \eta - \frac{1}{2c} N(\delta_t^i) \right]$$

which is monotonically increasing in δ_t^i when $\delta_t^i > 0$. As a result, $EUC(\delta_t^i) = 0$ defines a unique positive value $\hat{\delta}$ which we call the “churning threshold”. If $\delta_t^i > \hat{\delta}$ – optimism is above the churning threshold – then the principal finds it optimal to churn agents, while otherwise he will hire for the long term.

In either case, the best the principal can do is hire the agent who maximizes total surplus and create a contract so that the agent’s certainty equivalent is \bar{u} . If the principal can do this, then by offering one contract he can attract at least his desired agent. The principal’s utility

$$EU_p = (1 - B)\mu + \frac{1}{2c} B^2 - A$$

does not depend on who accepts the contract. So, conditional on the contract, the principal is indifferent to which agent he engages.

Whether the principal chooses to hire a new every period or keep one agent for the duration of the labor market, the principal benefits most at time 0 from hiring the agent with the most extremely positive beliefs: agent 1 (equations 3.25 and 3.26). Because this is the agent hired at time 0, we will refer to that as hiring agent δ_0^1 .

Proposition 9 *If $\delta_0^1 > \hat{\delta}$, then the principal chooses to hire a new agent every period.*

He will offer a contract

$$(A, B) = \left(\bar{u} - (1 + \delta_t^1 c) (\mu + \delta_t^1) - \frac{1}{2c} (1 + \delta_t^1 c)^2, 1 + \delta_t^1 c \right) \quad (3.29)$$

which agent (1, t) will accept each period

If $\delta_0^1 \leq \hat{\delta}$, then the agent will repeatedly hire the same agent, agent (1,0). The principal offers a contract

$$(A, B) = \left(\bar{u} - (1 + \delta_0^1 c) (\mu + \eta t + \delta_0^1) - \frac{1}{2c} (1 + \delta_0^1 c)^2, 1 + \delta_0^1 c \right) \quad (3.30)$$

which agent (1,0) will accept each period.

When the principal re-hires a given agent, the agent engages in two types of learning. First, his skills increase, resulting in a higher level of profitability for the project. This drives increases in project profitability. Second, the agent's beliefs converge – he remembers past events and updates. This means that, according to the principal, the agent requires a higher total contract value. In other words, memory and experience drive contract value.

In other models, the two gains from experience – skills and memory – are treated as one and the same, but the difference is important. Memory is a type of human capital whose value cannot be appropriated by the principal. The reason for this is that it does not enhance the value of the project; instead, it changes the valuation assigned by the agent. The result is that while the value of skills can be appropriated by an employer, the value of memory cannot.

3.4 Predictions and Discussion

3.4.1 Comparative Statics and Predictions

There are various proxies for measuring heterogeneous beliefs in the literature. The cleanest proxies can be obtained when expectations are measured directly. Diether, Malloy and Scherbina (2002) and Scherbina (2004) compute a measure of analyst dispersion forecast and link it to asset prices. Landier and Thesmar (2003) construct measures of biased beliefs of entrepreneurs from a survey, and find systematic differences in the behavior due to biased beliefs. Oberlechner and Osler (2003) construct

measures of overconfidence of currency traders by conducting a survey, and find that it helps internal promotion, but does not change profits.

Indirect measures can also be useful. In a seminal contribution, Harris and Raviv (1993) demonstrate that differences in beliefs generate volume, so, presumably, higher trading volume is correlated with more disperse beliefs within the marketplace, which in turn is correlated with more disperse beliefs within the firm or industry.

The most useful measures of heterogeneous beliefs at a firm or industry level are likely to be stock market trading volume and dispersion in analyst expectations. For our model, the most desirable data would be within-company data which would allow us to analyze individual contracts at middle and low levels of the hierarchy. This is likely to be hard to obtain.

The Cross Section of Incentive Contracts

If agents in the labor market pool and their potential principles have heterogeneous expectations about a given project, then agents with high δ will be available and will thus be hired. As a result, heterogeneous expectations will be associated with high powered incentives: high bonuses and relatively low salaries.

Prediction 1 [$\partial B/\partial\delta > 0$] *In the cross section of firms or industries, proxies for heterogeneous expectations will be positively correlated with high powered incentives.*

In order to test this prediction, proxies of heterogeneous beliefs could include firm-level and aggregate dispersion of analyst views and stock market turnover. High powered incentives can be measured with bonus clauses, option payments, and the like.

Industries will also differ to the extent that profit is enhanced by uncontractible effort. If this effort is relatively inexpensive, then contracts will have sharper incentives.

Prediction 2 [$\text{sign}(\partial B/\partial c) = \text{sign}(\delta)$] *In the cross section of industries, those with*

more observable effort or higher opportunity costs of effort will have sharper incentives.

When high powered incentives are driven by heterogeneous beliefs, they should also be associated with higher than average turnover for employees:

Prediction 3 *In a cross section of firms or industries, those with high powered incentives and heterogeneous expectations should have shorter employment duration and shorter term contracts.*

The Time-Series of Incentive Contracts

There is some evidence that the degree of heterogeneity of beliefs is time varying. In particular, Gilchrist, Himmelberg and Huberman (2003) report substantial time-variation in the aggregate level of analyst forecast dispersion. Intuitively, during boom times, agents may rely too heavily on recent events and as a result be over-optimistic regarding the near future. Our model predicts that incentives are steeper when over-confidence is higher. On the other hand, outside options (\bar{u}) and inherent project profitability affect only the salary, not the bonus portion of the optimal contract.

In particular, if agents overweight recent observations when they update their beliefs, then we expect them to be more optimistic during boom periods and less optimistic during recessions. As a result, we would expect to see higher powered incentives during boom times and not recessions.

Prediction 4 [$\partial B/\partial \delta > 0$] *Times with more heterogeneous expectations should be correlated with higher powered incentives. For example, the power of incentives should be correlated with the lagged business cycle.*

Prediction 5 *Times with high powered incentives and heterogeneous expectations should have shorter employment duration and shorter term contracts.*

These are time-series prediction that can be tested directly by computing the implied delta of stock option payments from Execucomp, and computing the aggregate degree of analyst dispersion from IBES, following Scherbina (2003).

3.4.2 Comparison to a Rational Model

One model that appears to generate similar predictions is a model of skill assessment.³ If agents have varying and unknown levels of skill, they may receive an uncertain signal about their ability. A principal could offer a high-powered contract, which only those with high signals would accept, and then fire all those who received a falsely positive signal. The result would be skill intensive industries would exhibit high powered incentives and labor market turnover while low skill intensity industries would not.

This type of theory predicts that at the industry level incentives and turnover would coexist. At the firm level, however, there would be a negative relationship: firms would use either more incentives or more turnover to assess skill, but they would have no need for both. Our model, by contrast, predicts that there should be a positive relationship between incentives and turnover at every level that sees heterogeneous expectations.

A skill search model also predicts that churning should stop after an agent's type has been fully revealed. Our model says that churning is a function of the market conditions, not a condition associated with any given agent, and so churning should continue. In other words, our model predicts that in industries or firms with high turnover, that turnover should be widespread and not just limited to recent hires.

Lastly, our business cycle prediction is very much at odds with the predictions that would be generated by a model in which agents had different skills. In such a model, different skills would come into play during good and bad economic times, meaning that churning would take place in the transitions between good and bad times. There would be little or no turnover, however, in more stable time periods. Our model predicts that churning will take place during good times and not bad times, rather than at the transitions.

³See, for example the career-concern model of Holmstrom (1999)

3.5 Conclusion

In this paper, we studied the incentive problems and labor market implications of a principle-agent model when agents' learning takes place in two dimensions. On the one hand, the skill level of agents increases over time. On the other hand, the differences in beliefs between the agent and the principle decreases as agents learn about the true probability distribution of their projects.

In an optimal contract, the principle exploits the heterogeneity of beliefs by “selling dreams” to the agent. However, over time, the difference of beliefs between the principle and the agent diminishes, at the same time as the skill level of the agent increases. When the heterogeneity in beliefs of newly hired agents is sufficiently large, our theory predicts a high degree of labor market churning, as the benefits from providing incentives for overoptimistic agents outweigh the costs of firing experienced ones. This is the main empirical prediction of the theory developed in the paper, and it fits well with the casual observation of the labor market of traders in financial institutions or the junior market for academics.

We derive a number of additional empirical predictions that differ from models with homogeneous priors. The slope of the incentive scheme is predicted to be increasing in the degree of heterogeneity of beliefs, as principles write contracts that exploit the possibility of side-bets. Industries with more high powered incentives are predicted to have a higher degree of churning. Larger firms have higher powered incentives, and the more so, the larger the differences in beliefs between principles and agents.

There are a number of promising avenues of future research. So far, we have derived the labor market implications when the principle is a monopolist. In our set-up, agents self-select once the principle offers a particular incentive contract. In a setting with several firms, we conjecture that firms will endogenously offer different contracts and thus attract different types of agents.⁴ In addition, heterogeneity of

⁴This idea has been proposed by Stephen Ross in a presentation of Ross (2004).

skill accumulation and the forecast error could be introduced.

Another promising research area is the endogenous choice of beliefs. Benabou and Tirole (2002) and Brunnermeier and Parker (2003) point out that choosing to be overconfident comes at the cost of making distorted choices. In our set-up, there is an additional trade-off: more overconfident agents are more likely to be churned and lose their job-specific skills due to unemployment. The existence of a churning equilibrium in the labor market will thus alter the optimal choice of overconfidence ex-ante, and might lead to time-inconsistency.

3.6 Appendix

3.6.1 A Discrete Time Analog

There is a single risky project in the economy which pays a dividend D_1 at time 1, with

$$D_1 \sim N(\mu + e, \sigma^2) \tag{3.31}$$

The rights to the project are owned by the principal, but to undertake the project he hires an agent. The agent exerts an unobservable level of effort, e . In choosing e , the agent affects the project's payoff.

In return for the agent's labor, the principal offers the agent a payment. The principal cannot observe e so he can make his payment contingent only on D_1 : $s(D_1)$. We further assume that the principal is restricted to offering linear contracts: contracts are of the form $s(D_1) = A + BD_1$.

Both the principal and agent are risk neutral over consumption, while the agent pays a quadratic cost for the effort (e) that he supplies the project.

We assume that the principal and agent use different probability measures, so their contract will reflect their differing expectations about the distribution of D_1 . The principal has probability measure P , and he believes that D_1 is distributed as in

equation 3.31. The principal receives expected utility equal to

$$E_0 U_p = E_0^P [D_1 - s(D_1)] \quad (3.32)$$

We have apparently given the principal objectively correct beliefs, but this is only because we wish to use the principal as our point of reference.

The agent, on the other hand, has probability measure Q under which

$$D_1 \sim^Q N(\mu + \delta + e, \sigma^2) \quad (3.33)$$

The agent receives expected utility from the contract equal to

$$EU_a = E_0^Q \left[s(D_1) - \frac{1}{2}ce^2 \right] \quad (3.34)$$

We further assume that the agent has an outside financial option worth \bar{u} .

Assuming he accepts the contract, the agent's problem is to find $\{e^*\}$ such that

$$e^* \in \operatorname{argmax} B(\mu + \delta + e) - \frac{1}{2}ce^2 \quad (3.35)$$

The principal's problem is to maximize total welfare subject to constraints:

$$\max_B \mu + e + B\delta - \frac{1}{2}ce^2 \quad (3.36)$$

s.t. (i) e^* solves the agent's problem

$$(ii) B^*(\mu + \delta + e^*) - \frac{1}{2}ce^{*2} + A^* \geq \bar{u}$$

From here we note that the agent will choose

$$e^* = \frac{B}{c}$$

and we pick up the text from equation 3.18.

3.6.2 Proofs

The following lemma is based on the appendix of Schattler and Sung (1993).

Lemma *Under standard integrability conditions, if*

$$V_t = \max_{\{e_t\}} \mathbf{E}_t \left[\int_t^1 z(s, D_s, e_s) ds \right]$$

exists, and

$$dD_t = (\mu + e_t) dt + \sigma dB_t$$

then V_t is such that

$$V_t = V_0 - \int_0^t z(s, D_s, e_s) ds + \int_0^t \nabla V_s \sigma dB_s$$

and

$$e_t^* = \operatorname{argmax}_{e_t} z(t, D_t, e_t) + e_t \nabla V_s$$

Proof: Define

$$M_t^e = \int_0^t z(s, D_s, e_s) ds + V_t$$

which represents the maximum value of $\mathbf{E}_0 \left[\int_0^1 z(s, D_s, e_s) ds \right]$ if some control e_s is used on $s \in [0, t]$ while the optimal control, e_s^* is used on $s \in (t, 1]$. Then

$$dM_t^e = z(t, D_t, e_t) dt + \nabla V_t dB_t^e$$

where the measure e is defined so that

$$dD_t = (\mu + e_t) dt + \sigma dB_t^e$$

Looking at two potential controls, u and e ,

$$V_t = M_t^e - \int_0^t z(s, D_s, e_s) ds = M_t^u - \int_0^t z(s, D_s, u_s) ds$$

so that

$$dM_t^u = dM_t^e - z(t, D_t, e_t) + z(t, D_t, u_t)$$

and

$$e_t dt + \sigma dB_t^e = u dt + \sigma dB_t^u$$

Because V_t is the value function and e_t^* is the optimal control, standard results tell us that $M_t^{e^*}$ is a martingale on $[0, 1]$. Thus M_t^e is a super martingale on $[0, 1]$. Moreover, M_t^e has a right-continuous version, and so we can decompose M_t^e into

$$M_t^e = N_t^e + L_t^e$$

where N_t^e is a martingale under measure P^e and L_t^e is a unique, predictable, integrable, increasing process. Let

$$N_t^e = \int_0^t \nabla V_t \sigma dB_t^e$$

so that

$$dM_t^u = \nabla V_t \sigma dB_t^e + dL_t^e - z(t, D_t, e, t) dt + z(t, D_t, u_t) dt$$

Equivalently, we can change measures to show

$$dM_t^u = \nabla V_t \sigma dB_t^u + dL_t^e - z(t, D_t, e, t) dt + z(t, D_t, u_t) dt + u \nabla V_t dt - e \nabla V_t dt$$

and

$$dM_t^u = \nabla V_t \sigma dB_t^u + dL_t^u$$

Together, these imply that

$$L_t^u = L_t^e + \int_0^t [z(s, D_s, u_s) + u_s \nabla V_s - z(s, D_s, e_s) - e_s \nabla V_s] ds$$

If $e = e^*$ is an optimal control, then $M_t^{e^*}$ is a martingale. Since $N_t^{e^*}$ is also a martingale and $L_0^{e^*} = 0$, then

$$L_t^{e^*} = 0$$

and so

$$L_t^u = \int_0^t [z(s, D_s, u_s) + u_s \nabla V_s - z(s, D_s, e_s^*) - e_s^* \nabla V_s] ds$$

Since L_t^u is strictly positive, e^* must maximize $z(s, D_s, e_s^*) + e_s^* \nabla V_s$, meaning $u = e^*$ is optimal when it sets $L_t^u = 0$. This proves the second statement of the lemma.

We can now reverse the definition of $M_t^{e^*}$ to show

$$V_t = M_t^{e^*} - \int_0^t z(s, D_s, e_s^*) ds$$

so that

$$dV_t = dM_t^{e^*} - z(s, D_s, e_s^*) ds = \nabla V_t dB^{e^*} - z(s, D_s, e_s^*) ds$$

which proves the first statement of the lemma.

3.6.3 The Kalman-Bucy Filter

This treatment of the Kalman-Bucy filter is taken from Liptser and Shiryaev (2000); a discrete time analogue can be found in Hamilton (1994).

The agent learns about the drift μ in the following process:

$$dD_t = [\phi(D_t, t) + \mu] dt + \sigma dB_t$$

where $\phi(D_t, t)$ is some function known to the agent. In our application, we have

$$\phi(D_t, t) = e_t^* = \frac{B}{c}$$

In most circumstances, learning and optimization can be separated, i.e. the agent does the filtering first, and then chosen the optimal effort under the filtered measure. In our application, however, the agent does not learn during the evolution of the

process – he applies at discrete intervals.

If we define

$$\begin{aligned} m_t &= \mathbf{E}_t [\mu] \\ \gamma_t &= \mathbf{E}_t [(\mu - m_t)^2] \end{aligned}$$

with initial beliefs m_0, γ_0 , then the Kalman-Bucy filter is

$$dm_t = \frac{\gamma_t}{\sigma^2} (dD_t - \phi(D_t, t) dt - m_t dt) \quad (3.37)$$

and the forecast error follows a Ricatti equation:

$$d\gamma_t = - \left(\frac{\gamma_t}{\sigma} \right)^2 dt$$

Solving this ODE for γ_t , we obtain:

$$\gamma_t = \frac{\sigma^2}{t + \sigma^2/\gamma_0}$$

We can now solve for the conditional expectation m_t . Rewrite the equation 3.37 as:

$$\frac{\gamma_0 t + \sigma^2}{\gamma_0 \sigma^2} dm_t + \frac{1}{\sigma^2} m_t dt = \frac{1}{\sigma^2} (dD_t - \phi(D_t, t))$$

or, after integrating the left hand side,

$$d \left(\frac{\gamma_0 t + \sigma^2}{\gamma_0 \sigma^2} m_t \right) = \frac{1}{\sigma^2} (dD_t - \phi(D_t, t))$$

Integrating again gives us:

$$m_t = \frac{\sigma^2 m_0 + \gamma_0 \int_0^t (dD_s - \phi(D_s, s))}{\gamma_0 t + \sigma^2} = \frac{\sigma^2 m_0 + \gamma_0 \left(D_t - D_0 - \int_0^t \phi(D_s, s) \right)}{\gamma_0 t + \sigma^2}$$

To finish, we need only substitute $\phi(D_t, t) = \frac{A}{c}$, and $\mu + \delta_t$ for m_t , and remember

that $D_t - D_0 = (\mu + A/c)t + \sigma B_t$ to find:

$$\delta_t = \frac{\sigma B_t + \delta_0 \sigma^2 / \gamma_0}{t + \sigma^2 / \gamma_0} \quad (3.38)$$

Bibliography

- [1] Benabou, R. and J. Tirole, 2002, Self-Confidence and Personal Motivation, *Quarterly Journal of Economics* 117/3, pp. 871-915.
- [2] Bertrand, M., and A. Schoar, 2003, Managing with Style: The Effect of Managers on Firm Policies, *Quarterly Journal of Economics* 118/4, pp. 1169-1208.
- [3] Bolton, P., J. Scheinkman, and W. Xiong, 2003, Executive Compensation and Short-termist Behavior in Speculative Markets, NBER Working Paper 9722.
- [4] Brunnermeier, M., and J. Parker, 2004, Optimal Expectations, mimeo, Princeton University.
- [5] Diether, K., C. Malloy, and A. Scherbina, 2002), Differences of Opinion and the Cross-Section of Stock Returns, *Journal of Finance* 57/5, pp. 2113-2141.
- [6] Gervais S., J. B. Heaton and T. Odean, 2003, Overconfidence, Investment Policy, and Executive Stock Options, mimeo, Fuqua School of Business and Haas School of Business.
- [7] Gervais, S., and T. Odean, 2001, Learning to be Overconfident, *Review of Financial Studies* 14/1, pp.1-27.
- [8] Grossman, Sanford J., and Oliver D. Hart, 1983, An Analysis of the Principal-Agent Problem, *Econometrica* 51/1, pp. 7-46.
- [9] Hamilton, J., 1994, Time Series Analysis, Princeton: Princeton University Press.

- [10] Harris, M. and A. Raviv, 1993, Differences of Opinion Make a Horse Race, *Review of Financial Studies* 6/3, pp. 473-506.
- [11] Harrison, J. M., and D. M. Kreps, 1978, Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations, *Quarterly Journal of Economics* 92/2, pp. 323-36.
- [12] Hart, O., and B. Holmstrom, 1987, The Theory of Contracts, in *Advances in Economic Theory, Fifth World Congress*, edited by T. Bewley, New York: Cambridge University Press.
- [13] Holmstrom, B., 1999, Managerial Incentive Problems - A Dynamic Perspective, *Review of Economic Studies* 66/1, pp. 169-182.
- [14] Holmstrom, B. and P. Milgrom, 1987, Aggregation and Linearity in the Provision of Intertemporal Incentives, *Econometrica* 55/2, pp. 303-328.
- [15] Kleiber, K. L., 2003, Managerial Compensation Contracts and Overconfidence, mimeo, WHU Otto Beisheim Graduate School of Management.
- [16] Kyle, A. S., and F. A. Wang, 1997, Speculation Duopoly with Agreement to Disagree: can Overconfidence Survive the Market Test?, *Journal of Finance* 52/5, 2073-2090.
- [17] Landier, A., and D. Thesmar, 2003, Contracting with Optimistic Entrepreneurs: Theory and Evidence, mimeo, Chicago GSB.
- [18] Liptser, R. S., and A. N. Shiryaev, 2000, *Statistics of Random Processes, Volume I & II*, Berlin: Springer-Verlag.
- [19] Malmendier, U., and G. Tate, 2003, CEO Overconfidence and Corporate Investment, mimeo, HBS and Stanford GSB.
- [20] Mann, S., and Locke, P., 2001, House Money and Overconfidence on the Trading Floor, mimeo, Neeley School of Business and George Washington University

- [21] Milgrom, P. and N. Stokey, 1982, Information, Trade and Common Knowledge, *Journal of Economic Theory* 26/1, pp. 17-27.
- [22] Miller, E. M., 1977, Risk, Uncertainty, and Divergence of Opinion, *Journal of Finance* 32, pp. 1151-68.
- [23] Osler, C., and T. Oberlechner, 2003, Overconfidence in Currency Markets, mimeo, Brandeis University.
- [24] Ross, S. A., 1973, The Economic Theory of Agency: The Principal's Problem, *American Economic Review* 63/2, pp. 134-39.
- [25] Ross, S. A., 2004, Compensation, Incentives, and the Duality of Risk Aversion and Riskiness, *Journal of Finance* 59/1, pp. 207-25.
- [26] Schattler, H. and Sung, J., 1993, The First-Order Approach to the Continuous-Time Principal-Agent Problem with Exponential Utility, *Journal of Economic Theory* 61/2, pp. 331-71.
- [27] Scheinkman, J., and W. Xiong, 2004, Overconfidence and Speculative Bubbles, *Journal of Political Economy* 111/6, pp. 1183-1219.
- [28] Scherbina, A., 2004, Analyst Disagreement, Forecast Bias and Stock Returns, mimeo, HBS.
- [29] Spence, M. and R. Zeckhauser, 1971, Insurance, Information, and Individual Action, *American Economic Review* (Papers and Proceedings) 61/2, pp. 380-7.
- [30] Wilson, R., 1969, The Structure for Incentives for Decentralization under Uncertainty, *La Decision* 171: Agregation et Dynamique des Ordres de Preference, G. Guilbaud (ed.), Paris: Centre National de la Recherche Scientifique, pp. 287-307.