

Forecasting the Forecasts of Others: Implications for Asset Pricing*

Igor Makarov[†] and Oleg Rytchkov
MIT Sloan School of Management

Job Market Paper

First Draft: November 2005

This Draft: January 2006

Abstract

We develop a dynamic asset pricing model with heterogeneously informed agents. Unlike previous research, we focus on the general case where differential information leads to the problem of “forecasting the forecasts of others” and to non-trivial dynamics of higher order expectations. In particular, we prove that the model does not admit a finite number of state variables. Using numerical analysis, we compare equilibria characterized by identical fundamentals but different information structure. We demonstrate that the distribution of information has substantial impact on equilibrium prices and returns. In particular, we show that asymmetric information can generate momentum in returns and high trading volume.

Keywords: asset pricing, asymmetric information, higher order expectations

JEL classification: D82, G12, G14

*We are very grateful to Leonid Kogan, Anna Pavlova, Steve Ross, and Jiang Wang for valuable discussions and suggestions. We thank participants at MIT Batterymarch seminar for helpful comments. All remaining errors are ours.

[†]Corresponding author. E-mail: imakarov@mit.edu. Phone: (617) 324-6709.

MIT Sloan School of Management, 50 Memorial Drive, E52-416, Cambridge, MA 02142.

1 Introduction

One of the major difficulties arising in the analysis of dynamic asset pricing models with asymmetric information is the problem of “forecasting the forecasts of others.” When investors possess different information about an asset’s payoff, prices generally reflect not only investors’ expectations of the asset’s fundamental value, but also their expectations of other investors’ expectations of it. Iterating this logic forward, prices must depend on the whole hierarchy of investors’ beliefs. This problem has interested economists for decades, as evidenced by Keynes’ (1936) much-cited comparison of financial markets with beauty contests: “We devote our intelligence to anticipating what average opinion expects the average opinion to be.”

In most cases the successive forecasts of the forecasts of others differ from one another. To account for all of them one needs an infinite number of variables, making the model not only analytically involved, but also numerically challenging. Almost all existing models get around this problem by employing various assumptions that, by restricting the possible dynamics of expectations, guarantee that all higher order beliefs can be described by a few carefully chosen state variables¹. It is not clear, however, that insights obtained from these models survive in a more general informational environment.

In this paper we do not impose standard simplifying restrictions and show that, as long as two fairly general conditions hold, the infinite regress problem cannot be avoided and an infinite number of state variables is required to describe the dynamics of prices². The two conditions are that each agent lack a component of fundamental information which is known to some other agents and that fundamentals evolve stochastically over time.

The first condition guarantees that information held by other agents is relevant to each agent’s payoff and, as a result, his beliefs about other agents’ beliefs affect his demand for the risky asset. We call this information setup differential and contrast it with the hierarchical setup, in which one agent is better informed than the other. In the latter case, the informed agent knows the forecasting error of the uninformed one and therefore does not need to forecast it, so higher order expectations collapse. The second condition forces agents to form new sets of higher order beliefs every period. Since no agent ever becomes fully informed,

¹See Brunnermeier (2001) for a review of asset pricing under asymmetric information.

²Surprisingly, this proof is not trivial and is not excessive. Despite the apparent simplicity of intuition leading to this conclusion, in each particular case it is not easy to show that the dynamics of the model do not admit description in terms of a finite number of state variables.

they all need to incorporate the entire history of prices into their predictions³.

By comparing the equilibria supported by the same fundamentals but with different distributions of information among investors, we are able to isolate the effect of information dispersion on expectations. We find that mistakes that agents make in forming their expectations are much larger under differential than under hierarchical information. Therefore, differential information gives rise to a larger deviation of prices from the benchmark case without information asymmetry. To better understand the dynamics of price, in addition to static contributions of expectations, we also study the joint dynamics of expectations and fundamentals. We show that the forecasting errors of agents are much more persistent when information is differentially distributed among agents, which is a direct consequence of the absence of superiorly informed agents who arbitrage these errors away.

Analysis of a more general informational setup allows us to evaluate the robustness of previous findings and get new insights. We find that under the differential information setup the absorption of information into prices can be very slow. As a result, returns can be positively autocorrelated, which may be a step towards an explanation of momentum. The driving force behind this effect in our model is underreaction of agents to new information. It should be emphasized that we consider an equilibrium model in which the diffusion of information into prices is an endogenous process: it is an equilibrium outcome of agents' portfolio decisions and the resulting effect on the price. This distinguishes our explanation of momentum from a number of behavioral theories which appeal to different cognitive biases⁴. Our model predicts that momentum is stronger in stocks with little analyst coverage, higher analysts dispersion forecasts, and higher volume, all of which are consistent with empirical evidence⁵.

Furthermore, our model allows us to investigate the effects of information dispersion on trading volume, the empirically observed high levels of which present a puzzle to financial economists. We differentiate between two types of trades: informational trades between informed agents and exogenous trades with liquidity traders. We show that under the hierarchical setup there is almost no trade between informed and uninformed agents. This

³This observation suggests that price histories may play an important role in financial markets in which asymmetric information is ubiquitous, thus lending support to technical analysis, which is often employed in practice.

⁴See Barberis, Shleifer, and Vishny (1998), Daniel, Hirshleifer, and Subrahmanyam (1998), Hong and Stein (1999) among others.

⁵See Hong, Lim, and Stein (2000), Lee and Swaminathan (2000), Verardo (2005).

is intuitive, since the uninformed are aware of their disadvantage and therefore averse to trade. As a result, in this framework trading volume is almost exclusively determined by the properties of the exogenously assumed process for noise trader demands. In contrast, in a framework with differential information, trading volume between informed agents is high, since no one has a clear advantage. More importantly, we demonstrate that the contribution of informational trade to total volume can be significantly higher than that of exogenous volume.

There is a vast literature related to dynamic asset pricing with asymmetric information. In most papers, however, the role of higher order expectations is limited. Grundy and McNichols (1989), and Brown and Jennings (1989) study two-period models, which are very restrictive for analysis of dynamic effects of differential information. Singleton (1987), and Grundy and Kim (2002) consider models in which all private information becomes public after one or two periods. As a result, investors' learning problem in these papers becomes a static one, weakening the effects of asymmetric information on expectations and prices.

Enabling private information to be long-lived allows for non-trivial interplay between expectations and fundamentals which sometimes reverts the conclusions of simplified models. For example, in contrast to Grundy and Kim (2002), we demonstrate that volatility of returns under differential information may be lower than in an otherwise identical economy with no information asymmetry.

Wang (1993, 1994) and He and Wang (1995) analyze multi-period economies with asymmetric information. To avoid the aforementioned infinite regress problem Wang (1993, 1994) considers a model with hierarchical information. In the model of He and Wang (1995) agents have differential information about the unknown value of payoff, which, however, does not evolve over time. Our setup naturally extends these models by allowing both a general information structure and stochastic evolution of fundamentals.

The above papers assume competitive markets. Admati and Pfleiderer (1988), Foster and Viswanathan (1996), Back, Cao, and Willard (2000), and Bernhardt, Seiler, and Taub (2004), among others, consider dynamic noncompetitive models under asymmetric information. The theme of our paper is also aligned with another strand of literature which explicitly analyzes higher order expectations. Allen, Morris, and Shin (2004) argue that under asymmetric information agents tend to underreact to private information, making price biased towards the public signal. Bacchetta and Wincoop (2004) show that under asymmetric information price deviations from its fundamental value can be large. Having a fully-fledged dynamic

model enables us to provide a more thorough analysis of agents' expectations and their dynamics as well as to give specific predictions about their relationship to the behavior of prices, returns and trading volume.

The rest of the paper is organized as follows. Section 2 describes the model. In Section 3 we solve for the equilibria in benchmark cases of full and hierarchical information dispersion setups. In Section 4 we consider differential information. In Sections 5 and 6 we analyze higher order expectations and their impact on prices and returns. We compare variances and autocorrelations of returns under different information structures and argue that the distribution of information among agents can substantially affect the observed time series characteristics. Section 7 is devoted to analysis of trading volume generated in our model. In Section 8 we demonstrate that differential information accompanied by evolving fundamentals can generate momentum in returns. Section 9 concludes. Technical details are presented in Appendices A, B, C and D.

2 The Model

In this section, we present a noisy rational expectation model with two assets. The first asset is a riskless asset in perfectly elastic supply that generates a rate of return $1 + r$. The second asset is a claim on a hypothetical firm which pays no dividends⁶ but has a chance of being liquidated every period. We assume that the probability of liquidation in period $t + 1$, given that the firm has survived until period t , is equal to λ . Upon liquidation the firm pays its equity holders a stochastic liquidation value V_t . This liquidation value can be decomposed into two components: $V_t = V_t^1 + V_t^2$, and each component evolves according to a first-order autoregressive process:

$$V_{t+1}^j = aV_t^j + b_V\epsilon_{t+1}^j, \quad j = 1, 2.$$

We assume that $\epsilon_t^j \sim \mathcal{N}(0, 1)$ are *i.i.d.* across time and components. For simplicity we take identical parameters a and b_V for the processes V_t^1 and V_t^2 . The total amount of risky equity⁷ available to rational agents is $1 + \theta_t$, where $\theta_t \equiv b_\Theta\epsilon_t^\Theta$ and $\epsilon_t^\Theta \sim \mathcal{N}(0, 1)$.

There is an infinite set of competitive rational investors indexed by i and uniformly distributed on a unit interval $[0, 1]$. We assume that investors are fully rational and know the

⁶We model the firm as not paying dividends for simplicity, since the current dividend would be an additional signal about future cash flows.

⁷This can be interpreted as supply of stock by noise traders. Following Grossman and Stiglitz (1980), we introduce stochastic amount of equity to prevent prices from being fully revealing.

model. Each of them is endowed with some piece of information about the fundamentals V_t^1 and V_t^2 . We assume that investors are mean-variance optimizers and each investor i submits the demand X^i which is proportional to his expectation of excess stock return Q_{t+1} :

$$X_t^i = \frac{1}{\alpha} \frac{E[Q_{t+1}|\mathcal{F}_t^i]}{\text{Var}[Q_{t+1}|\mathcal{F}_t^i]}, \quad Q_{t+1} = \lambda V_{t+1} + (1 - \lambda)P_{t+1} - (1 + r)P_t. \quad (1)$$

Here \mathcal{F}_t^i is the information set of investor i at time t . All investors are assumed to have the same coefficient of risk aversion α . We focus on a rational expectation equilibrium of this model which is defined by two conditions:

- 1) all agents rationally form their demands according to (1);
- 2) market clearing condition holds: $\int X_t^i di = 1 + \theta_t$.

Before we turn to analysis of equilibrium, it is worthwhile to make several comments about the model. First of all, we make the model very stylized, since we want to demonstrate and analyze the “forecasting the forecasts of others” problem in the simplest setting. In particular, we assume that all shocks are normally distributed and this property is inherited by other random variables in the model, leading to the linear form of conditional expectations and, therefore, to a linear equilibrium. Next, we consider a model with an infinite horizon and focus on stationary equilibria which enables us to use powerful methods from the theory of stationary Gaussian processes. Finally, a major simplification is achieved by assigning agents’ mean-variance preferences. This assumption is similar to the assumption of logarithmic utility with lognormally distributed shocks in that hedging demand is zero. Since calculation of hedging demand in the economy with infinite number of state variables is complicated by itself⁸, sidestepping this problem allows us to preserve tractability of the model but still relate equilibrium price to agents’ higher order beliefs and characterize their dynamics.

In the most general case, information sets of investors \mathcal{F}_t^i are different, investors have to forecast the forecasts of others, and non-trivial higher order expectations appear. Since they are responsible for the major properties of the equilibrium, we analyze their contribution in detail. As a basis for our subsequent analysis, it is useful to represent the price in terms of fundamentals and expectations of agents. It is convenient to first define the *weighted average expectation operator* $\bar{E}_t^w[x]$ of agents as follows:

$$\bar{E}_t^w[x] = \int \frac{\omega_i}{\Omega} E[x|\mathcal{F}_t^i] di, \quad \Omega = \int \omega_i di, \quad \omega_i = \frac{1}{\alpha} \frac{1}{\text{Var}[Q_{t+1}|\mathcal{F}_t^i]}$$

Note that the weights ω_i are endogenous and determined by the conditional variances of excess returns given investors’ information sets. The expectations of agents with better information

⁸See Schroder and Skiadas (1999) for some results in this case.

get larger weights than those who are less informed. Using the market clearing condition we can derive a relation between the current price and the next period price:

$$P_t = -\frac{1 + \theta_t}{\Omega(1 + r)} + \frac{1}{1 + r} \bar{E}_t^w [\lambda V_{t+1} + (1 - \lambda) P_{t+1}].$$

Iterating this relation forward and imposing the no-bubble condition, we get

$$P_t = -\frac{1}{\Omega(r + \lambda)} - \frac{1}{\Omega(1 + r)} \theta_t + \frac{a\lambda}{1 + r} \sum_{s=0}^{\infty} \left(\frac{1 - \lambda}{1 + r} \right)^s \bar{E}_t^w \bar{E}_{t+1}^w \dots \bar{E}_{t+s}^w V_{t+s}. \quad (2)$$

This equation represents the price as a series over iterated weighted average expectations of future values of V_t : we have arrived at a mathematical formulation of forecasting the forecasts of others. It highlights two essential difficulties. The first is that the law of iterated expectations need not hold because agents may have different information; this point was recently emphasized by Allen, Morris and Shin (2004). The second and even more significant obstacle is that the current price also depends on agents' future expectations which, in turn, depend on future prices. Consequently, in order to compute their expectations, we have to solve for the entire sequence of prices as a fixed point. Since this problem is quite complicated, before attempting to find a solution for the general case, let us first consider some special cases in which the solution is not as involved.

3 Benchmarks: full information and hierarchical information equilibria

As a starting point, we consider the full information setup, which serves as a benchmark for our further analysis. Full information means that all investors $i \in [0, 1]$ observe both components V_t^1 and V_t^2 and their information sets are

$$\mathcal{F}_t^i = \{P_\tau, V_\tau^1, V_\tau^2 : \tau \leq t\}.$$

In this case we are back to the representative agent framework, and the law of iterated expectations holds: $\bar{E}_t^w \bar{E}_{t+1}^w \dots \bar{E}_{t+s}^w V_{t+s} = E_t V_{t+s} = a^s V_t$. Now observing the price is sufficient to infer the demand of noise traders θ_t . We have the following proposition:

Proposition 1. Suppose that

- 1) all investors observe V_t ;
- 2) $2\sqrt{2}b_V b_\Theta \frac{\lambda(1-\lambda)}{1+r-a(1-\lambda)} \leq \frac{1}{\alpha}$.

Then there exists a full information equilibrium in which the equilibrium price of the risky asset is given by

$$P_t = -\frac{1}{\Omega(r + \lambda)} + \frac{a\lambda}{1 + r - a(1 - \lambda)}V_t - \frac{1}{\Omega(1 + r)}\theta_t. \quad (3)$$

$$\Omega = \frac{(1 + r - a(1 - \lambda))^2}{4b_V^2\lambda^2(1 + r)^2} \left(\frac{1}{\alpha} + \sqrt{\frac{1}{\alpha^2} - \frac{8b_V^2b_\Theta^2\lambda^2(1 - \lambda)^2}{(1 + r - a(1 - \lambda))^2}} \right) \quad (4)$$

Proof. See Appendix A.

The obtained price function has a structure which is common to linear rational expectations models⁹. The first term corresponds to a risk premium for uncertain payoffs. The second term is the value of expected future payoffs discounted at the risk-free rate adjusted for the probability of liquidation. The third term compensates the investors for noise trading related risk.

Formally, the equations determining equilibrium price admit two solutions. One of them is given in Proposition 1, and we take this solution as the full information benchmark in the future. The reason for discriminating between equilibria is that the other solution is unstable, meaning that minor errors in agents' behavior significantly impact prices and destabilize the economy. Having this in mind, we consider only the full information equilibrium which is most sensible from the economic point of view.

Now consider the equilibrium with hierarchical information¹⁰, which means that investors can be ranked according to the amount of their information: some investors are better informed than others. Formally, the information sets of investors at time t are hierarchically embedded in each other and generate a filtration: $\mathcal{F}_t^i \subseteq \mathcal{F}_t^{i'} \subseteq \dots$. We focus on the simplest case, and assume that there are only two types of investors which we denote as 1 and 2. Investors of type 1, which are indexed by $i \in [0, \gamma]$, are informed and observe both V_t^1 and V_t^2 . Investors of type 2, with $i \in (\gamma, 1]$, are partially informed and observe V_t^2 only. We can write their information sets of informed and uninformed investors as

$$\mathcal{F}_t^1 = \{P_\tau, V_\tau^1, V_\tau^2 : \tau \leq t\}, \quad \mathcal{F}_t^2 = \{P_\tau, V_\tau^2 : \tau \leq t\}.$$

There are several reasons why this informational structure is interesting. First of all, it is an intermediate setup between the full information and the differential information equilibria.

⁹See Campbell and Kyle (1993), Wang (1993)

¹⁰The idea to analyze hierarchical information setup in order to avoid the infinite regress problem was suggested by Townsend (1983) and elaborated in the asset pricing context by Wang (1993, 1994).

Despite the investors having heterogeneous information, the infinite regress problem does not arise and we can find a closed-form solution. The intuition behind this result is simple and can be easily conveyed in terms of expectations. When trying to extract the unknown piece of information from the price, investors of type 2 form their expectations $\hat{V}_t^1 = E[V_t^1|\mathcal{F}_t^2]$ about the current value of V_t^1 . Since all agents of type 2 make an identical estimation error, \hat{V}_t^1 is a new state variable influencing the price of the asset. In their turn, the investors of type 1 need to form their own expectations about expectations of type 2 investors, $E[\hat{V}_t^1|\mathcal{F}_t^1]$, and in the general case of differential information, it would be represented by another state variable. However, since $\mathcal{F}_t^2 \subseteq \mathcal{F}_t^1$ we get $E[\hat{V}_t^1|\mathcal{F}_t^1] = E[E[V_t^1|\mathcal{F}_t^2]|\mathcal{F}_t^1] = \hat{V}_t^1$ and the infinite regress problem does not arise. Basically, since the type 1 agents have all the information, they can, without mistake, deduce the mistake of type 2 agents, thus their prediction of the price is accurate. So the hierarchical information case illustrates how iterated expectations collapse and the state space of the model remains finite dimensional.

The hierarchical information equilibrium in our model is characterized by Proposition 2.

Proposition 2. If investors of type 1, with $i \in [0, \gamma]$, observe V_t^1 and V_t^2 and investors of type 2, with $i \in (\gamma, 1]$, observe only V_t^2 the equilibrium price of the risky asset is given by

$$P_t = -\frac{1}{\Omega(r + \lambda)} + p_V V_t + p_\Theta \theta_t + p_\Delta (\hat{V}_t^1 - V_t^1), \quad (5)$$

where p_V , p_Θ , p_Δ and Ω are constants which solve a system of nonlinear equations given in Appendix B.

Proof. See Appendix B.

4 Differential information equilibrium

Now consider the informational structure in which all agents are endowed only with a piece of relevant information and the rest of the information is never revealed. Again, assume that there are two types of agents, $j = 1, 2$ with $i \in [0, \gamma]$ and $i \in (\gamma, 1]$ respectively, such that their information sets are given by

$$\mathcal{F}_t^1 = \{P_\tau, V_\tau^1 : \tau \leq t\}, \quad \mathcal{F}_t^2 = \{P_\tau, V_\tau^2 : \tau \leq t\}. \quad (6)$$

It means that the agents of type j can observe only V^j and the history of prices. Let us show how the problem of “forecasting the forecasts of others” arises in this case. First of all, due to the presence of noise traders, the price is not fully revealing, i.e. knowing the price

and their own component of information V^j , the agents cannot infer the other component V^{-j} . However, the information about V^{-j} is relevant to agent j , since it helps him predict his own future payoff and, consequently, to form his demand for the asset. Moreover, due to the market clearing condition, the information of each investor is partially incorporated in the price, each agent has an incentive to extract the missing information of the other type from the price. Therefore, an agent will form his own expectations about the unknown piece of information. For example, agent 1 forms his expectations about agent 2's information. These expectations of agent 1 affect his demand and, subsequently, the price. So the inference problem of agent 2 is not only to extract the information of agent 1, but also the expectations of agent 1 about the information of agent 2. Agent 1, in turn, faces a similar problem; we can see how the infinite regress starts to appear.

The above reasoning might seem to be quite general, however, it does not always produce an infinite set of different higher order expectations. He and Wang (1995) provide an example how the higher order expectations can be reduced to first-order expectations even when investors have differential information. They consider a similar setup but assume that the firm is liquidated with probability one at some future time T and that the liquidation value does not evolve over time. In this situation, investors also try to predict the weighted average of investors' expectations \widehat{V} of V . The paper demonstrates that \widehat{V} can be written as a weighted average of V conditional on public information (price) and the true value of V . Given this, investor i 's expectation of \widehat{V} is a weighed average of his first-order expectations, conditional on price and on his private signals. Averaging them, one can show that second-order expectations of V can be again expressed as weighted average of V conditional on price and the true value of V . As will be shown later, this logic breaks down when V evolves stochastically over time.

It is necessary to distinguish between the cases with finite vs. infinite dimensional state space because they are conceptually different and call for different solution techniques. In the former case, the major problem is to find appropriate state space variables. In the latter, the search for a finite set of state variables that can capture the dynamics is worthless by default, and the solution of such models presents a greater challenge.

To provide the ground for rigorous treatment of the “forecasting the forecasts of others”, we introduce the concept of Markovian dynamics. Let $(\Omega, \mathcal{F}_t, \mu)$, $t \in \mathbb{Z}$ be a complete probability space equipped with a filtration \mathcal{F}_t . In what follows, all the processes are assumed to be defined on this space.

Definition. Let X_t be an adaptive random process. We say that X_t admits Markovian dynamics if there exists a collection of $n < \infty$ adaptive random processes $\bar{Y}_t = \{Y_t^i\}$, $i = 1..n$, such that the joint process (X_t, \bar{Y}_t) is Markov, that is

$$Prob(X_t \leq x, \bar{Y}_t \leq y | X_\tau, \bar{Y}_\tau : \tau \leq t-1) = Prob(X_t \leq x, \bar{Y}_t \leq y | X_{t-1}, \bar{Y}_{t-1})$$

Obviously any Markov process admits Markovian dynamics. The next example will further help to clarify the ideas.

Example. Let ε_t , $t \in \mathbb{Z}$ be *i.i.d.* standard normal random variables. Define $X_t = \varepsilon_t - \theta\varepsilon_{t-1}$, an MA(1) process. X_t is not a Markov process, or even an n-Markov process: $Prob(X_t | X_\tau : \tau \leq t-1) \neq Prob(X_t | X_{t-1}, \dots, X_{t-n})$ for any n . However, X_t can be easily extended to a Markov process if one augments it with ε_t .

An important consequence of X_t admitting Markovian dynamics is that the filtered process \hat{X}_t then also admits Markovian dynamics, provided that signals obey the Markov property. As a result, all relevant information is summarized by a finite number of variables.

Applying the concept of Markovian dynamics to our model we get the following result.

Proposition 3. Suppose agents' information sets are given by

$$\mathcal{F}_t^j = \{P_\tau, V_\tau^j : \tau \leq t\}, \quad j = 1, 2.$$

Then in the linear equilibrium of the described economy the price process does not admit Markovian dynamics.

Proof. See Appendix C.

Although we give a detailed proof in Appendix C, it is useful to make some comments on it here. The idea behind the proof is to use the following result from the theory of stationary Gaussian processes: if the process admits Markovian dynamics, then it is described by a rational function in the frequency domain. We start with the assumption that the price admits Markovian dynamics. The main part of the proof is to show that it is impossible to satisfy the market clearing condition and to simultaneously solve the optimal filtering problem of each agent working only with rational functions. This contradiction proves that the equilibrium price does not admit Markovian dynamics and the infinite regress problem is there.

To highlight the significance of this result from the theoretical standpoint, we refer to the paper by Townsend (1983), which inspired the study of the infinite regress problem and coined the term “forecasting the forecasts of others”. Townsend attempted to create a setup

in which traders would have to estimate the beliefs of others in order to solve their own forecasting problems. However, Sargent (1991) and Kasa (2000) show how to reduce all higher order expectations in his model to just a small number of cleverly chosen low order expectations. Since then, a lot of effort has been made to state the necessary and sufficient conditions for the infinite regress problem to exist. We demonstrate that our setup is, in a sense, a minimal model where this phenomenon appears. We know from the result of He and Wang (1995) that if the value of the payoff remains constant over time, it is possible to reduce higher order expectations to first order expectations. In our model, we relax just this condition. It is still interesting to search for other cases, in which solution can take a simple form. Our result, however, severely restricts the set of possible candidates. It suggests that the infinite regress problem is almost unavoidable if one is willing to consider a situation more general than ones previously studied.

The result also provides support for technical analysis. The simplicity of the fundamentals in our model leads to a straightforward solution in the case of complete information. However, asymmetric information results in highly non-trivial price dynamics. Now, to be as efficient as possible, agents have to use the entire price history in their predictions: as stated in Proposition 3, they cannot choose a finite number of state variables to summarize the price dynamics. This suggests that in financial markets, where fundamentals are not as simple and asymmetric information is commonplace, price history may be informative for investors.

Unfortunately, systems with an infinite number of state variables are very difficult to analyze and, in general, do not admit an analytical solution. So, in our solution we have to rely on reasonable numerical approximations, the details of which are relegated to Appendix D.

5 Analysis of higher order expectations

When information is dispersed among agents, higher order expectations play an important role in price formation. Moreover, higher order expectations determine not only the wedge between the price and the fundamental value of the firm, but also the statistical properties of prices and returns. Thus, analysis of higher order expectations and especially of their dynamical properties is helpful for understanding the impact of information distribution on prices and returns.

It is convenient to decompose the price as given in equation (2) into the part determined

by fundamentals and corrections arising as a consequence of heterogenous expectations:

$$P_t = -\frac{1}{\Omega(r + \lambda)} + \frac{a\lambda}{1 + r - a(1 - \lambda)}V_t - \frac{1}{\Omega(1 + r)}\theta_t + \frac{a\lambda}{1 + r} \sum_{s=0}^{\infty} \left(\frac{1 - \lambda}{1 + r}\right)^s (\bar{E}_t^w \bar{E}_{t+1}^w \dots \bar{E}_{t+s}^w - E_t)V_{t+s}, \quad (7)$$

where E_t is the expectation operator with respect to full information. Here the differences $\Delta_t^s = (\bar{E}_t^w \bar{E}_{t+1}^w \dots \bar{E}_{t+s}^w - E_t)V_{t+s}$ represent pure effects of asymmetric information. The price decomposition (7) is valid for any information distribution. So we can apply it to both the hierarchical and differential information cases and compare contributions of higher order expectations in each.

In the hierarchical information case, all terms in the infinite series of expectations can be calculated explicitly. In particular, simple calculation yields

$$\Delta_t^s = a^s \frac{\omega_2}{\Omega} \frac{1 - (c \frac{\omega_1}{\Omega})^{s+1}}{1 - c \frac{\omega_1}{\Omega}} \left(\hat{V}_t^1 - V_t^1 \right).$$

Here c , ω_1 and ω_2 are constants defined in Appendix B. As expected, all higher order expectations terms are proportional to the estimation error $\hat{V}_t^1 - V_t^1$. It means that all terms in the infinite series are perfectly correlated and the series collapses to one term $p_{\Delta}(\hat{V}_t^1 - V_t^1)$, greatly simplifying the analysis.

We have already shown that with differential information all higher order expectations are different so we can calculate Δ_t^s only numerically. In setting parameter values for our numerical solution, we assume that the length of one period is a month. It is reasonable to set the probability of liquidation λ to 5% annually, so that the expected life of a firm is 20 years. We make risk-free rate r equal to 1% annually. We let α , the coefficient of risk aversion, equal 3, which is a commonly chosen value, for example, as in Campbell, Grossman, and Wang (1993). We set the mean-reversion parameter a to 0.85. We make the size of supply shocks b_{Θ} equal to 15%. In Section 7 we relate this parameter to volume turnover. In order for risk premium and return volatility to roughly match their empirical counterparts, we set b_V equal to 1.2. With these parameters we get risk premium equal to slightly more than 7% and return volatility to about 15% annualized.

A natural starting point for our analysis is to quantify the contribution of Δ_t^s to the price in each period. To do this, we first compute unconditional variances of Δ_t^s for $s = 0 \dots 20$. The results are presented in Figure 1. We observe that variances of Δ_t^s are significantly greater

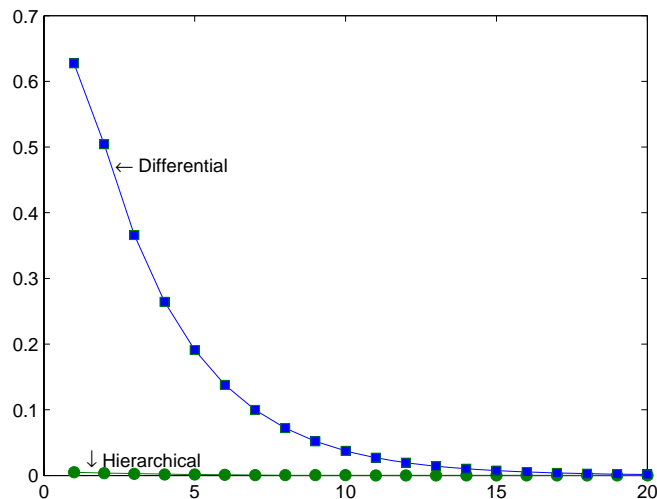


Figure 1: Variances of Δ_t^s for $s = 0.20$ in the hierarchical and differential information equilibria.

s	0	1	2	3
0	1	0.99	0.99	0.99
1	0.99	1	0.99	0.99
2	0.99	0.99	1	0.99
3	0.99	0.99	0.99	1

Table 1: Correlations of $\Delta_t^s = \bar{E}_t^w \bar{E}_{t+1}^w \dots \bar{E}_{t+s}^w V_{t+s} - E_t V_{t+s}$, $s = 1, 2, 3$ when investors are differentially informed.

than zero for a large number of values of s in the differential information case, as opposed to those in the hierarchical case, which suggests a greater effect of information asymmetry in the former.

To confirm the presence of this effect, we investigate the correlations between Δ_t^s for different s . If Δ_t^s are not correlated or negatively correlated, their contributions might cancel, reducing their aggregate impact on the price. On the other hand, if Δ_t^s are highly correlated then it might be possible to identify a small number of them which capture all dynamics.

From Table 1 we observe that Δ_t^s , indeed have extremely high correlations. It follows that if we want to approximately characterize the direction of price deviation from the full information benchmark, it is enough to consider the stochastic behavior of Δ_t^0 , which is a cross-sectional average of estimation errors. In general, the exact degree of correlation

	Hierarchical	Differential
$\text{Corr}(V_t, \Delta_t^0)$	-0.02	-0.26
$\text{Corr}(\theta_t, \Delta_t^0)$	-0.99	-0.71
$\text{Corr}(\Delta_{t-1}^0, \Delta_t^0)$	0.004	0.65

Table 2: Correlations of $\Delta_t^0 = \bar{E}_t^w V_t - E_t^w V_t$ with V_t and θ_t in the hierarchical and differential information equilibria.

depends on parameter values, and is inversely related to the total amount of noise in the system.

To get better understanding of the effects of higher order expectations we compute correlations between Δ_t^0 and the fundamentals V_t and θ_t . Since all Δ_t^s are highly correlated with each other, the correlations between Δ_t^s , $s \geq 1$ and the fundamentals are almost identical and we do not report them.

From Table 2 we can see that Δ_t^0 is negatively correlated with contemporaneous values of both V_t and θ_t , which admits a simple intuitive explanation. Let us start with the negative correlation with V_t . When V_t is high, the difference $\Delta_t^0 = \bar{E}_t V_t - V_t$ is low since at least some investors do not know exactly the value of V_t and their average estimation is biased towards the mean value of V_t , which is 0. The intuition behind negative correlation of Δ_t^0 and θ_t is also straightforward. If there is a positive supply shock, the price of the asset goes down. However, some investors cannot perfectly distinguish this shock from a negative shock to V_t , and therefore their estimation of V_t is again low.

This intuition is valid for both hierarchical and differential information, but the numbers in these cases are significantly different. We see that in the case of hierarchical information, informed investors take most of the fundamental risk, but leave some of the liquidity risk to the uninformed. With differential information, no agents are perfectly informed. It translates into a greater average mistake about the fundamentals.

It is also interesting to examine the autocorrelation of Δ_t^s . We see that there is almost no autocorrelation in the hierarchical case, and a pretty high value 0.65 under the differential information. The reason for this is that under hierarchical information, fully informed investors can take advantage of the mistakes of the uninformed and therefore arbitrage them away. When investors are differentially informed they all make errors. Moreover, the errors made by one type depend not only on fundamentals but also the errors made by another

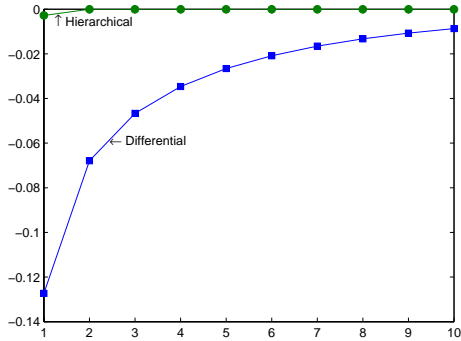


Figure 2: Impulse response of Δ_t^0 to the shock ϵ_t^1 in the models with different informational structures.

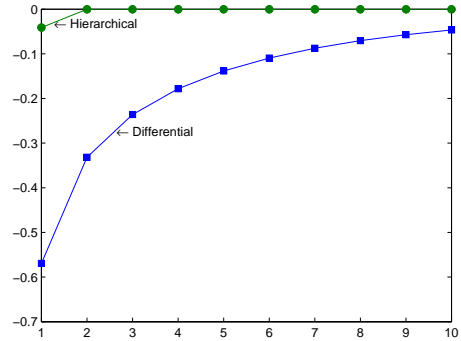


Figure 3: Impulse response of Δ_t^0 to the shock ϵ_t^Θ in the models with different informational structures.

type of investors. Without fully informed arbitrageurs, mistakes are much more persistent in comparison with hierarchical information case since it takes much longer to correct them.

We conclude our study of Δ_t^0 by depicting how it depends on a particular shock represented by its coefficients in the decomposition over the current and past shocks under different information structures.

Figures 2 and 3 provide more support for the above results. We observe that in the differential information case, Δ_t^0 not only has much higher negative loadings on both fundamentals and supply shocks than in the hierarchical one, but also its response to shocks declines significantly more slowly.

6 Basic properties of prices and returns

Having studied stochastic behavior of higher order expectations, let us now consider how they affect the basic properties of prices and returns. The results are given in Table 3.

We see that under full information returns are negatively auto-correlated. Intuition for this fact is quite simple: under full information, the expected return and price are negatively correlated. Negative supply shock in period t makes price go up in period t . Since the shock is transitory, the next period price is expected to be lower, thus leading to negative auto-correlation. We see even more negative auto-correlation with the hierarchical information setup. However, in the case of differential information the first order autocorrelation is positive and is about half percent. This suggests that asymmetric information might lead to momentum in stock returns. We postpone the detailed discussion of its relation to momentum

	Full	Hierarchical	Differential
	(%)	(%)	(%)
$E(Q_t)$	7.56	7.57	7.16
$\text{Std}(Q_t)$	15.88	15.89	15.47
$\text{Corr}(Q_t, Q_{t-1})$	-0.04	-0.06	0.57
$\text{Var}(P_t)$	25.6	25.6	24.7

Table 3: Risk premium and statistical properties of price and excess return in models with different informational structures.

for a while and first give the intuition for the observed pattern.

Consider an increase in price from the perspective of any investor. If this increase cannot be explained by his own information about fundamentals, the price can be high for two reasons: either there was a positive shock to the liquidation value observed by other investors, or there was a negative supply shock. These shocks have different impacts on returns. Consider first the situation in which the price increase is caused by a positive shock to the liquidation value. An uninformed agent, seeing the price increase, may perceive it as a negative supply shock and, therefore, underreact to it. Since V^j is a persistent process, the next period price will be higher than the agent expects, causing him to update his expectations, thus creating continuations in returns. Consider now the situation in which there is a negative supply shock. Again, seeing the price going up, the uninformed agent confuses it with a positive shock to the liquidation value. But since noise is *i.i.d.*, its effect is mostly temporary, and the price in the next period will tend to be lower than he expects, leading to reversals in returns. Which effect dominates depends on the relative sizes of these two shocks and on how quickly information impounds into prices. The longer it takes for shocks to liquidation value to get into prices, the larger the induced positive autocorrelation of returns. We have already seen that under differential information mistakes are much more persistent in comparison with hierarchical information, which explains autocorrelation behavior in these cases.

We see that in the hierarchical information case volatility is slightly larger than under full information, and volatility is lowest under differential information. This observation contradicts the conclusion of Grundy and Kim (2002), who assert that differential information causes returns to be more volatile than in the benchmark case with no information asymmetry. The cause for this discrepancy is that in Grundy and Kim's model private information is short

lived, so investors can only trade on their information for one period, and therefore trade more aggressively. If, on the other hand, information is not revealed every period, as in our model, investors have plenty of time to trade on their information. As a result, it takes a long time for shocks to be impounded into prices, making returns less volatile.

Because we assigned our agents a mean-variance demand over a one period horizon, the volatility of one period returns has a direct effect on their perception of risk, producing an inverse relation between expected returns and volatility. This is a result of our simplifying assumptions, and a more thorough modeling of agents' preferences, for example as in Wang (1993) would be required if one is interested in rigorous analysis of the effect of asymmetric information on risk premium.

It is interesting to consider the price variance. From Table 3 we can see that the price variance is almost identical in full information and hierarchical information cases, but goes down in the differential information setup. From our previous analysis (cf. Table 1), we can see that the higher order expectations have two opposite effects on the price volatility. On one hand, they represent additional volatile state variables the inclusion of which increases total price volatility. On the other hand, these state variables are negatively correlated with V_t , and this correlation is higher for differential information. This leads to a decrease in volatility. The overall effect depends on which of the two effects dominates. For the given choice of parameters, these effects almost cancel each other in the hierarchical information case, and the second effect dominates in the case of differential information, in which price volatility goes down.

7 Trading volume

In this section we examine the basic properties of trading volume under different information dispersion setups. This question has received a significant amount of attention in the past. For example, Wang (1994) conducts an extensive analysis of stock trading volume under hierarchical information and He and Wang (1995) study it under differential information. However, as we have already pointed out, these papers employ various simplifying assumptions to avoid the infinite regress problem. As a result, there is no easy way to compare the findings among those models. In our work both hierarchical and differential information are nested within the same model, which makes it possible to conduct such analysis.

Let us first give a definition of volume in our model. Since the average number of shares

	I(%)	II(%)	III(%)
Vol ¹	16.9	33.7	37.8
Vol ²	16.9	1.2	37.8
Vol ¹²	0	0.2	10.5
Vol ^{NT}	16.9	16.9	16.9
Vol ^{Tot}	16.9	17.1	27.4
Corr(ΔP_t , Vol _t ^{Tot})	0.1	0.7	6.7

Table 4: Normalized trading volume in models with different informational structures. Vol^{*i*} – average trading volume of type *i* investors, Vol¹² – trading volume between two classes of investors, Vol^{NT} – trading volume of noise traders, Vol^{Tot} – aggregate trading volume. I – full information equilibrium, II – hierarchical information equilibrium, III – differential information equilibrium.

in our model is equal to one, what we refer to as trading volume is actually the turnover. If in period *t* agent *i* holds X_t^i shares, but in period *t* + 1 his holdings are X_{t+1}^i then his (unsigned) trading volume is $|X_{t+1}^i - X_t^i|$. We are interested in average trading volume $Vol^i = E|X_{t+1}^i - X_t^i|$ of each agent and the relation between volume and information the particular agent has. All liquidity shocks, produced by noise traders, are absorbed by rational investors, and their trading volume is mostly determined by these shocks. However, when investors have different information, they also trade with each other and, on average, this volume can be characterized as $Vol^{12} = E[\int |X_{t+1}^i - X_t^i| di - |\theta_{t+1} - \theta_t|] / 2$. Because this volume is generated endogenously, we call it informational volume. On the other hand, the trading volume of noise traders $Vol^{NT} = E|\theta_{t+1} - \theta_t|$ is completely exogenous in the model. The results are collected in Table 4.

Under full information, the volume is completely exogenous: no trades occur between the informed agents. They simply absorb liquidity shocks, equally splitting the volume. In the situation with hierarchical information, the informed agents absorb most of the trades, since the uninformed agents are aware of their disadvantage and are therefore averse to trade. Their volume is not zero because they try to trade against the noise traders, but they occasionally make mistakes and end up trading against informed investors. In the case of differential information, the situation is very different. Agents of different types are not afraid of trading against one another, which leads to a high trading volume between them, as well as total

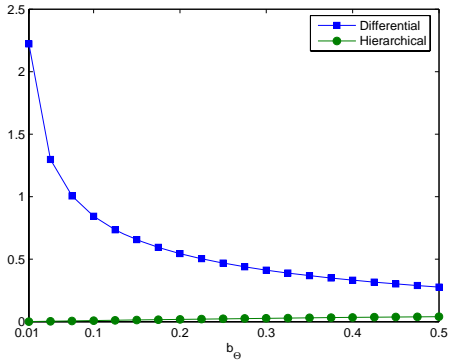


Figure 4: The ratio of informational volume in the hierarchical and differential information equilibria to the exogenous noise trading volume $\text{Vol}^{I2}/\text{Vol}^{NT}$ as a function of noise trading intensity b_{Θ} .

volume.

It is interesting to consider the behavior of trading volume with respect to the amount of noise trading. Figures 4 and 5 show the ratio of the informational volume to the volume of noise traders and total volume, respectively, for both hierarchical and differential information.

We observe that total volume is increasing under both setups, which is driven primarily by the increase in the exogenous volume of liquidity traders. With the normality assumption about the underlying shocks, we have

$$E|\theta_{t+1} - \theta_t| = \frac{2b_{\Theta}}{\sqrt{\pi}} \quad (8)$$

which is linear in b_{Θ} . It is more instructive to consider the behavior of the ratio. We can see that it displays a very different behavior. In the case of hierarchical information, the ratio is increasing in b_{Θ} . The increase in the level of noise trading represents a better trading opportunity for the uninformed traders, so they start to trade more. However, the more they trade, the more often they trade against the informed investors. In the case of differential information the only obstacle to trade is the no-trade theorem. Price becomes more and more informative as the level of noise trading decreases. But as the analysis suggests, in this case, investors, although trade less in absolute terms, trade much more relative to the trades by liquidity traders. This result suggests that asymmetric, especially differential, information can help explain high trading volume levels observed in financial markets.

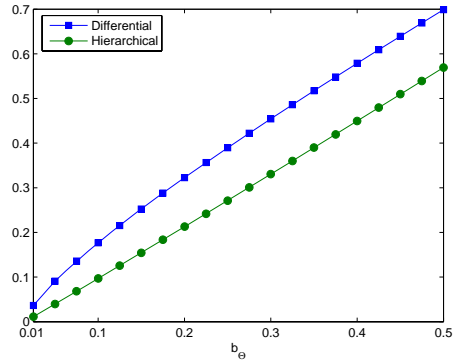


Figure 5: Total trading volume Vol^{Tot} in the hierarchical information and differential information equilibria as a function of noise trading intensity b_{Θ} .

We can also notice that the model is capable of producing another stylized fact about volume: the positive correlation of trading volume with absolute price changes. From Table 4, we see that correlation increases from full information to hierarchical information, and is strongest under differential information. In the case of full information, the price moves whenever any shock occurs. However, change in volume is only caused by supply shocks. As we move from full information to hierarchical and to differential, more and more trades come from shocks to fundamentals, which results in increased correlation.

8 Momentum

In a seminal paper by Jegadeesh and Titman (1993) it is shown that buying past winners and shorting past losers and holding them over 3 to 12 months generates high abnormal profits. Since its discovery, momentum has been one of the most resilient anomalies that challenge the market efficiency hypothesis. Despite a vast empirical literature about momentum¹¹, there are few theories that try to explain it. These theories are traditionally classified into rational and behavioral.

The rational theories provide risk-based explanations of momentum relating momentum to systematic risk of cash flows. In Berk, Green, and Naik (1999) momentum results from the slow evolution of the project portfolio of the firm. Johnson (2002) demonstrates that momentum can arise in a fully rational and complete information setting with stochastic expected dividend growth rates. However, analysis in both papers is conducted in the partial equilibrium framework, with an exogenously specified pricing kernel.

Other researchers have turned to behavioral models, which generally attribute momentum to underreaction or delayed overreaction, caused by cognitive biases. In Barberis, Shleifer and Vishny (1998) investors, due to the conservative bias, tend to underweight new information when they update their priors. In Daniel, Hirshleifer and Subrahmanyam (1998) investors are overconfident and overestimate the precision of their signals. As a result, they overreact to private information, but not to public information. Hong and Stein (1999) assume that information is slowly revealed to “news-watchers,” who observe future payoff relevant signals but do not use price as a source of information.

In general, there can be three possible sources of momentum profits¹². The first is that

¹¹Jegadeesh and Titman (2005) is a recent review.

¹²See Lo and MacKinlay (1990).

winners tend to be stocks with high unconditional expected returns¹³. Second, if one assumes that factors are positively autocorrelated, then winners could be stocks with high loadings to these factors. Finally, it might come from positive autocorrelation of idiosyncratic returns. In all but the first explanation, some components of stock returns should be positively autocorrelated. Therefore, any theory aimed to explain momentum should be able to generate positive autocorrelations in returns.

Results of Section 6 demonstrate that excess returns are positively autocorrelated under differential information. When investors form their expectations of Q_{t+1} at time t , however, they do not observe Q_t , but rather the history of prices. If liquidation does not occur at time t , the realized excess return is $\Delta P_t^e = P_t - (1+r)P_{t-1}$, and it is this return that we would need to see positively correlated with Q_{t+1} to observe momentum. This correlation in our base model is negative, since if θ_t are *i.i.d.* the incentive to follow the contrarian strategy is very strong. The logic behind positive autocorrelation of Q_t suggests, however, that if we reduce this incentive by modifying the process for θ_t we will be able to make correlation of Q_{t+1} with ΔP_t^e positive. Further we demonstrate that this is indeed the case, but first it is instructive to show why it is very difficult to generate momentum with θ_t being *i.i.d.*

Let \tilde{X}_t^i be the demeaned demand of the investor i . The market clearing condition implies

$$\int \tilde{X}_t^i di = \theta_t. \quad (9)$$

Recall that $X_t^i = \omega_i E[Q_{t+1} | \mathcal{F}_t^i]$. If we multiply both sides of equation (9) by ΔP_t^e and take the unconditional expectation, then by the law of iterated expectations, we arrive at the following equation:

$$\text{Cov}(Q_{t+1}, \Delta P_t^e) = \frac{1}{\Omega} E(\theta_t \Delta P_t^e). \quad (10)$$

If correlation Q_{t+1} with ΔP_t^e is positive, then when agents see the price increase they have higher expected returns, and therefore should hold a larger number of shares. If θ_t are *i.i.d.*, however, this is highly unlikely, since $E(\theta_t \Delta P_t^e) = E(\theta_t P_t)$. This quantity is negative in most models, because a positive supply shock normally leads to lower price.

In deriving this result, we use the fact that agents have myopic preferences. In general, there will also be a hedging demand. Note, however, that if the hedging demand results solely from information asymmetry, then it is a linear combination of agents' forecasting mistakes, and therefore is orthogonal to the public information set. Since everyone observes the price, the covariance of the hedging demand with ΔP_t^e is zero, which leaves the left hand side of

¹³Jegadeesh and Titman (2002) present evidence against this explanation.

equation (10) unchanged. As a result, the distribution of information between agents can change the magnitude of the correlation but not the sign. If θ_t follows an $AR(1)$ process, $\theta_t = a_\Theta \theta_{t-1} + b_\Theta \varepsilon_t^\Theta$, a simple calculation yields

$$E(\theta_t \Delta P_t^e) = \frac{b_\Theta^2}{1 - a_\Theta^2} (1 - a_\Theta(1 + r)) \sum_{k=0}^{\infty} p_k^\theta a_\Theta^k, \quad (11)$$

where $p_0^\theta, p_1^\theta, \dots$ are coefficients in the decomposition of the price over $\theta_t, \theta_{t-1}, \dots$. The sum $\sum_{k=0}^{\infty} p_k^\theta a_\Theta^k$ is likely to be negative in most models, since it means that price is negatively affected by shocks ε_t^Θ . Therefore, the sign of the correlation Q_{t+1} with ΔP_t^e again depends only on the sign of $(1 - a_\Theta(1 + r))$ and not on information dispersion or any other model parameters¹⁴. It is worthwhile to compare this result with that of Brown and Jennings (1989), who are able to generate positive autocorrelation for a wider range of parameters in a two period, but otherwise similar model. This difference underscores the importance of considering a stationary economy where the initial conditions have little effect on properties of equilibrium.

Next, we demonstrate that if we allow θ_t to have more general dynamics, information dispersion has a qualitative effect on prices. Having a more sophisticated process for θ_t obviously makes the model ad hoc. But the assumption of θ_t being *i.i.d.* or following $AR(1)$ process is also quite arbitrary. The only reason we allow liquidity traders to exist in our model is to break down the no-trade theorem and make informational effects possible. While for most of our results modeling θ_t as *i.i.d.* random variables seems to be the most innocuous, in the study of autocorrelation patterns, as we have shown, it is quite restrictive. To keep the model as parsimonious as possible, we consider the case in which θ_t follows an $AR(2)$ process, $\theta_t = a_1 \theta_{t-1} + a_2 \theta_{t-2} + b_\Theta \varepsilon_t^\Theta$.

For our numerical analysis we set a_1 and a_2 to 1.43 and -0.48 , respectively. This choice is somewhat arbitrary. It guarantees that correlations of Q_{t+1} with ΔP_t^e in the full information case are negative at all lags, but at the same time, makes the incentive to trade against liquidity traders small enough. We verify that similar results can be attained at other parameter values as well. Table 5 presents the correlation of Q_{t+1} with ΔP_t^e in models with different informational structures. We can see that these values, although smaller in magnitude, resemble the pattern of the autocorrelations in Table 3. The intuition is also similar: now, when we sufficiently reduce the impact of liquidity traders, the effect of slow diffusion of shocks to fundamentals dominates, thus leading to momentum in the differential information

¹⁴Wang (1993) illustrates this observation.

	Full	Hierarchical	Differential
	(%)	(%)	(%)
$\text{Corr}(Q_t, \Delta P_t^e)$	-0.02	-0.03	0.01

Table 5: Correlation Q_{t+1} with ΔP_t^e in models with different informational structures.

case. It should be emphasized that the diffusion of information is an endogenous process in our model and is consistent with demands of fully rational investors. It distinguishes our results from those of Hong and Stein (1999), who take the slow rate of information revelation as an assumption. Of course, since we are looking at just one stock, this is not the whole story about momentum: we do not take into account diversification at the limit, but this is beyond the scope of this paper.

There are several empirical regularities which support the informational explanation of momentum and which are consistent with our model. Hong, Lim, and Stein (2000) show that momentum predominantly resides in small stocks, and that, controlling for size, momentum is greater for firms with little analyst coverage. These stocks are less informationally transparent, and if momentum is really due to slow diffusion of information into prices then exactly these stocks should exhibit the strongest momentum behavior. Verardo (2005) finds that momentum is more pronounced in stocks with high dispersion of analysts' forecasts. This observation is also consistent with the suggested information theory of momentum. Indeed, if analysts have diverse opinions on a particular stock it is likely that it is more difficult to get objective and reliable information about the firm. Hence, it takes more time for news to be incorporated into prices. As a result, in accordance with our theory, this stock is more prone to momentum.

Although our model suggests that momentum arises due to slow diffusion of information, it lacks a well defined parameter controlling the precision of information that agents have. So, to get sharper predictions consistent with the empirical facts, we consider an extension of our model. We still assume that there are two types of investors, $j = 1, 2$ and investors of type j know V_τ^j , $\tau \leq t$. But now we introduce a third component V_t^3 , which is observed by both types of investors, so that the total value of the firm consists of three parts: $V_t = V_t^1 + V_t^2 + V_t^3$. Again, V_t^3 follows an AR(1) process, $V_{t+1}^3 = aV_t^3 + b_V^* \varepsilon_{t+1}$.

The third component allows us to control the magnitude of the information dispersion. To separate the impact of information from the effect of changing fundamentals we keep the

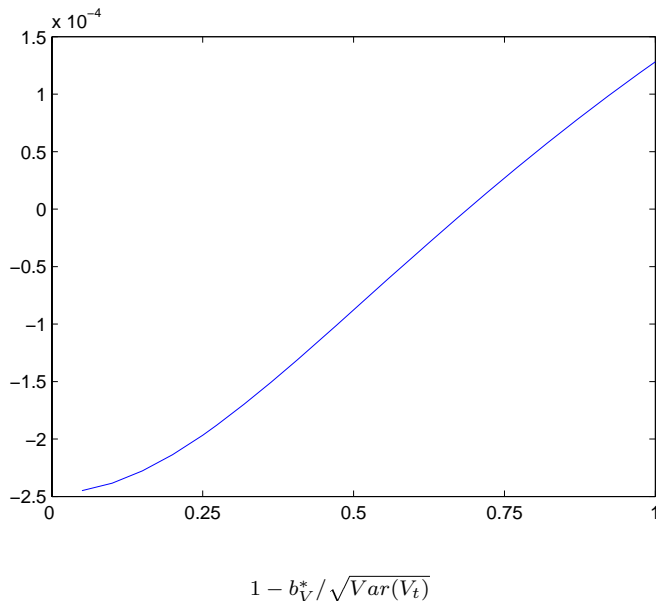


Figure 6: Correlation of Q_{t+1} with ΔP_t^e as a function of the information dispersion measured by $1 - b_V^*/\sqrt{Var(V_t)}$.

variance of V_t constant. By increasing the contribution of V_t^3 to the total firm value V_t and decreasing that of V_t^1 and V_t^2 , we decrease the information dispersion among agents. To make the results comparable across the sections we choose $Var(V_t) = 2 \times 1.2^2 = 2.88$ and control the contribution of V_t^3 by means of b_V^* . Thus, if b_V^* is close to zero the contribution of V_t^3 is negligible and we arrive at the differential information case with maximum information dispersion. On the contrary, if b_V^* is close to $\sqrt{2.88}$ the third component dominates and we get the full information case with zero information dispersion. We measure how diverse opinions are among the agents by $1 - b_V^*/\sqrt{Var(V_t)}$.

To gain a better understanding of the relation between momentum and information dispersion we plot the correlation of Q_{t+1} with ΔP_t^e as a function of $1 - b_V^*/\sqrt{Var(V_t)}$ in Figure 6. We see that correlation increases monotonically with information dispersion and eventually converges to the positive correlation observed under differential information. This observation is consistent with the results of Verardo (2005), thus providing support to our information-based theory of momentum.

Our model is also consistent with the empirical relation between momentum and volume reported in Lee and Swaminathan (2000) who find that momentum is higher in stocks with higher trading volume. Indeed, we demonstrated earlier that in our model the trading volume is an increasing function of the intensity of noise trading b_Θ . Figure 7 shows the behavior of

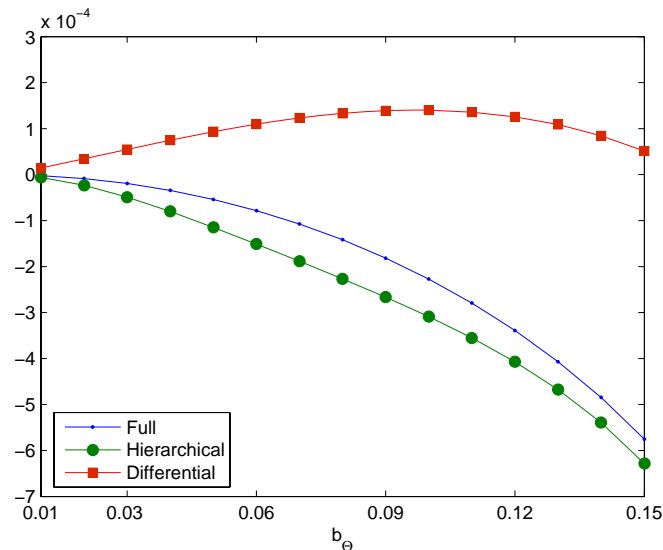


Figure 7: Correlation Q_{t+1} with ΔP_t^e as a function of the noise trading intensity b_{Θ} in the full, hierarchical, and differential information equilibria.

the correlation of Q_{t+1} with ΔP_t^e as a function of b_{Θ} .

We can observe that for differential information the correlation first increases and then decreases. The intuition behind this result can be explained as follows. Recall that positive correlation in our model is due to underreaction to fundamentals. At small values of b_{Θ} its increase leads to a slower diffusion of information, which dominates the direct effect of stronger supply shocks responsible for negative correlation. However, as b_{Θ} becomes larger, the direct effect dominates, producing a decline in correlation. Empirically we know that monthly turnover for most stocks is within 20%¹⁵, which implies, according to Figure 5, that volatility of θ_t is less than 15%. Since now θ_t follows a more persistent process, this suggests a value of b_{Θ} less than 5%. So we would expect most stocks to be on the increasing part of Figure 7, providing a positive cross-sectional relation between momentum and trading volume.

9 Concluding remarks

This paper presents a dynamic equilibrium model of asset pricing under different information dispersion setups. The model allows us to clarify the mechanics behind the infinite regress problem and explicitly demonstrate the effect of information distribution. By analyzing differential information coupled with time evolving fundamentals we are able to provide new

¹⁵See, for example, Lo and Wang (2000).

insights about the behavior of prices, returns, and trading volumes.

Due to the complexity of the problem, we made a number of simplifying assumptions. It is reasonable to believe that the intuition we gain from our analysis can be applied to more realistic models as well. There are several directions in which our paper can be developed. First, it would be interesting to consider a setup with multiple stocks and analyze the effect of information distribution on cross-correlations of prices and returns¹⁶. Next, we consider myopic investors who do not have hedging demand, significantly simplifying the model, since otherwise we would have to solve a dynamic program with an infinite dimensional space of state variables. The impact of hedging could be non-trivial and needs further research.

In our model the agents are exogenously endowed with their information and can neither buy new information, nor release their own information if they find this exchange profitable. It might be interesting to relax this assumption and to introduce the market for information. This direction was explored in a static setting by Verrecchia (1982), Admati and Pfleiderer (1986), and others but dynamic properties of the market for information are not thoroughly explored¹⁷.

Although our analysis pertains mostly to asset pricing, the insights about various aspects of the “forecasting the forecasts of others” problem and iterated expectations, as well as the intuition behind our results, are much more general and also relevant for other fields. For example, higher order expectations naturally arise in different macroeconomic settings (Woodford (2002)), in the analysis of exchange rate dynamics (Bacchetta and Wincoop (2003)), in models of industrial organization where, for example, firms have to extract information about unknown cost structure of competitors (Vives (1988)). The application of our approach and analysis of higher order expectations in these fields might be fruitful and need further research.

¹⁶See Admati (1985), Easley and O’Hara (2004), and Hughes, Liu, and Liu (2005), among others, for a static analysis.

¹⁷See Naik (1997b) for analysis of monopolistic information market in a dynamic framework.

Appendix A

Proof of Proposition 1.

Our starting point is a representation of equilibrium price (7). If all investors know V_t^1 and V_t^2 then the infinite sum collapses to zero and we get

$$P_t = -\frac{1}{\Omega(r + \lambda)} + \frac{a\lambda}{1 + r - a(1 - \lambda)}V_t - \frac{1}{\Omega(1 + r)}\theta_t.$$

So the only remaining problem is to calculate Ω which is endogenous and is determined by conditional variance of Q_{t+1} . Simple calculation yields

$$\text{Var}(Q_{t+1}|\mathcal{F}_t^i) = \frac{2\lambda^2(1+r)^2b_V^2}{(1+r-a(1-\lambda))^2} + \frac{(1-\lambda)^2b_\Theta^2}{\Omega^2(1+r)^2}.$$

By definition of Ω

$$\Omega = \int \frac{1}{\alpha} \frac{di}{\text{Var}[Q_{t+1}|\mathcal{F}_t^i]} = \frac{1}{\alpha} \frac{1}{\text{Var}[Q_{t+1}|\mathcal{F}_t^i]},$$

so we get an equation for Ω

$$\frac{1}{\alpha} \frac{1}{\Omega} = \frac{2\lambda^2(1+r)^2b_V^2}{(1+r-a(1-\lambda))^2} + \frac{(1-\lambda)^2b_\Theta^2}{\Omega^2(1+r)^2}$$

or, equivalently,

$$\frac{2\lambda^2(1+r)^2b_V^2}{(1+r-a(1-\lambda))^2}\Omega^2 - \frac{\Omega}{\alpha} + \frac{(1-\lambda)^2b_\Theta^2}{(1+r)^2} = 0.$$

It is a quadratic equation which has real solutions only if its discriminant is non-negative.

This condition can be rewritten as

$$2\sqrt{2}b_Vb_\Theta \frac{\lambda(1-\lambda)}{1+r-a(1-\lambda)} \leq \frac{1}{\alpha}.$$

Under this condition there is a full information solution with Ω given in Proposition 1.

Appendix B

Proof of Proposition 2.

If investors are hierarchically informed the infinite sequence of iterated expectations collapses to one term $\hat{V}_t^1 = E[V_t^1|\mathcal{F}_t^2]$, which is a new state variable of the economy. So, making conjecture that the price is a linear function of state variables we get :

$$P_t = p_0 + p_{V^1}V_t^1 + p_{V^2}V_t^2 + p_\Theta\theta_t + p_\Delta(\hat{V}_t^1 - V_t^1),$$

where $p_0, p_{V^1}, p_{V^2}, p_\Theta$ and p_Δ are constants. For future convenience we isolate the difference $\hat{V}_t^1 - V_t^1$ which is an estimation error of uninformed investors. The dynamics of \hat{V}_t^1 can be

found from the filtering problem of uninformed agents. To solve this problem we use the following theorem¹⁸.

Theorem (Kalman - Bucy filter).

Let the discrete linear system be described as

$$x_t = \Phi x_{t-1} + \Gamma \epsilon_{x,t}, \quad y_t = M x_t + \epsilon_{y,t},$$

where x_t is the n -vector of unobservable state variables at t , y_t is the m -vector of observations at t . Φ , Γ and M are $(n \times n)$, $(n \times r)$, $(m \times n)$ constant matrices, respectively. $\epsilon_{x,t}$ and $\epsilon_{y,t}$ are r -vector and m -vector white Gaussian sequences: $\epsilon_{x,t} \sim \mathcal{N}(0, Q)$, $\epsilon_{y,t} \sim \mathcal{N}(0, R)$. $\epsilon_{x,t}$ and $\epsilon_{y,t}$ are independent. Denote the optimal estimation of x_t at time t as \hat{x}_t :

$$\hat{x}_t = E[x_t | y_\tau : \tau \leq t]$$

and define

$$\Sigma = E[(x_t - \hat{x}_t)(x_t - \hat{x}_t)' | y_\tau : \tau \leq t].$$

Then

$$\hat{x}_t = (I_n - KM)\Phi\hat{x}_{t-1} + Ky_t,$$

$$\Sigma = (I_n - KM)(\Phi\Sigma\Phi' + \Gamma Q\Gamma'),$$

$$K = (\Phi\Sigma\Phi' + \Gamma Q\Gamma')M'[M(\Phi\Sigma\Phi' + \Gamma Q\Gamma')M' + R]^{-1},$$

where I_n is $(n \times n)$ identity matrix.

In our case the system of unobservable state variables is $V_{t+1}^1 = aV_t^1 + b_V\epsilon_{t+1}^1$ and the partially informed investors effectively observe $Z_t = (p_{V^1} - p_\Delta)V_t^1 + p_\Theta\theta_t$. So we have the following mapping: $x_t = V_t^1$, $y_t = Z_t$, $\Phi = a$, $\Gamma = b_V$, $M = p_{V^1} - p_\Delta$, $R = (p_\Theta b_\Theta)^2$, $Q = 1$. Applying the above theorem we get the solution to the filtering problem of partially informed investors and the dynamics of the new state variable \hat{V}_t^1 :

$$\hat{V}_t^1 = a(1 - k(p_{V^1} - p_\Delta))\hat{V}_{t-1}^1 + k(p_{V^1} - p_\Delta)V_t^1 + kp_\Theta\theta_t, \quad (12)$$

where k solves the following quadratic equation:

$$p_\Theta^2 b_\Theta^2 a^2 (p_{V^1} - p_\Delta) k^2 + (p_\Theta^2 b_\Theta^2 (1 - a^2) + b_V^2 (p_{V^1} - p_\Delta)^2) k - b_V^2 (p_{V^1} - p_\Delta) = 0. \quad (13)$$

¹⁸See Jazwinski (1970) for textbook discussion of linear filtering theory.

Equation (12) leads to the following AR(1) dynamics of the estimation error:

$$\hat{V}_t^1 - V_t^1 = ac(\hat{V}_{t-1}^1 - V_{t-1}^1) - b_V c \epsilon_t^1 + kb_\Theta p_\Theta \epsilon_t^\Theta,$$

where $c = 1 - k(p_{V^1} - p_\Delta)$.

Next, consider the demand functions of investors and the market clearing condition. Aggregate demand of partially informed investors is

$$\begin{aligned} X_t^2 &= (1 - \gamma) \frac{E[Q_{t+1} | \mathcal{F}_t^2]}{\alpha \text{Var}[Q_{t+1} | \mathcal{F}_t^2]} \\ &= \omega_2 \left((1 - \lambda)p_0 + a(\lambda + (1 - \lambda)p_{V^2})V_t^2 + a(\lambda + (1 - \lambda)p_{V^1})\hat{V}_t^1 - (1 + r)P_t \right) \\ &= \omega_2 \left((1 - \lambda)p_0 + a(\lambda + (1 - \lambda)p_{V^1})V_t^1 + a(\lambda + (1 - \lambda)p_{V^2})V_t^2 + a(\lambda + (1 - \lambda)p_{V^1})(\hat{V}_t^1 - V_t^1) - (1 + r)P_t \right), \end{aligned}$$

where we use our conjecture for the price function. By definition, $\omega_2 = (1 - \gamma) / (\alpha \text{Var}[Q_{t+1} | \mathcal{F}_t^2])$.

Similarly, aggregate demand of informed investors is:

$$X_t^1 = \gamma \frac{E[Q_{t+1} | \mathcal{F}_t^1]}{\alpha \text{Var}[Q_{t+1} | \mathcal{F}_t^1]} = \omega_1 \left(a\lambda V_t + (1 - \lambda)E[P_{t+1} | \mathcal{F}_t^1] - (1 + r)P_t \right)$$

or, using the law of motion of $\hat{V}_t^1 - V_t^1$

$$X_t^1 = \omega_1 \left((1 - \lambda)p_0 + a(\lambda + (1 - \lambda)p_{V^1})V_t^1 + a(\lambda + (1 - \lambda)p_{V^2})V_t^2 + acp_\Delta(1 - \lambda)(\hat{V}_t^1 - V_t^1) - (1 + r)P_t \right),$$

where $\omega_1 = \gamma / (\alpha \text{Var}[Q_{t+1} | \mathcal{F}_t^1])$. Market clearing condition $X_t^1 + X_t^2 = 1 + \theta_t$ gives

$$\begin{aligned} P_t &= \frac{\Omega(1 - \lambda)p_0 - 1}{\Omega(1 + r)} + \frac{a(\lambda + (1 - \lambda)p_{V^1})}{1 + r}V_t^1 + \frac{a(\lambda + (1 - \lambda)p_{V^2})}{1 + r}V_t^2 \\ &\quad - \frac{1}{\Omega(1 + r)}\theta_t + \frac{a(\omega_2\lambda + (1 - \lambda)(\omega_2p_{V^1} + \omega_1cp_\Delta))}{\Omega(1 + r)}(\hat{V}_t^1 - V_t^1), \end{aligned} \quad (14)$$

where $\Omega = \omega_1 + \omega_2$. Comparing (14) with the conjectured equation for price we get a set of equations for the coefficients p_0 , p_{V^1} , p_{V^2} , p_Θ and p_Δ :

$$\begin{aligned} p_0 &= \frac{\Omega(1 - \lambda)p_0 - 1}{\Omega(1 + r)}, & p_{V^1} &= \frac{a(\lambda + (1 - \lambda)p_{V^1})}{1 + r}, \\ p_{V^2} &= \frac{a(\lambda + (1 - \lambda)p_{V^2})}{1 + r}, & p_\Theta &= -\frac{1}{\Omega(1 + r)}, \\ p_\Delta &= \frac{a(\omega_2\lambda + (1 - \lambda)(\omega_2p_{V^1} + \omega_1cp_\Delta))}{\Omega(1 + r)}. \end{aligned}$$

These equations can be resolved as:

$$p_0 = -\frac{1}{\Omega(r + \lambda)}, \quad p_{V^1} = p_{V^2} = \frac{a\lambda}{1 + r - a(1 - \lambda)},$$

$$p_{\Theta} = -\frac{1}{\Omega(1+r)}, \quad p_{\Delta} = \frac{\omega_2 \lambda a(1+r)}{(1+r-a(1-\lambda))(\Omega(1+r) - \omega_1 a c(1-\lambda))}.$$

Thus we obtain coefficients p_{V^1} and p_{V^2} in terms of exogenous parameters of the model. In order to get p_0 , p_{Θ} and p_{Δ} we have to calculate $\text{Var}[Q_{t+1}|\mathcal{F}_t^1]$, $\text{Var}[Q_{t+1}|\mathcal{F}_t^2]$ and supplement our equations by definitions of ω_1 and ω_2 . Indeed,

$$\begin{aligned} \text{Var}[Q_{t+1}|\mathcal{F}_t^1] &= \text{Var}[b_V(\lambda + (1-\lambda)p_{V^1})(\epsilon_{t+1}^1 + \epsilon_{t+1}^2) + b_{\Theta}(1-\lambda)p_{\Theta}\epsilon_{t+1}^{\Theta} \\ &\quad + (1-\lambda)p_{\Delta}(-b_V c \epsilon_{t+1}^1 + k b_{\Theta} p_{\Theta} \epsilon_{t+1}^{\Theta})] \\ &= b_V^2 [(\lambda + (1-\lambda)(p_{V^1} - c p_{\Delta}))^2 + (\lambda + (1-\lambda)p_{V^2})^2] + b_{\Theta}^2(1-\lambda)^2 p_{\Theta}^2(1 + k p_{\Delta})^2, \end{aligned}$$

$$\begin{aligned} \text{Var}[Q_{t+1}|\mathcal{F}_t^2] &= \text{Var}[b_V(\lambda + (1-\lambda)p_{V^1})(\epsilon_{t+1}^1 + \epsilon_{t+1}^2) + (\lambda + (1-\lambda)p_{V^1})aV_t^1 + b_{\Theta}(1-\lambda)p_{\Theta}\epsilon_{t+1}^{\Theta} \\ &\quad + (1-\lambda)p_{\Delta}(ac(\hat{V}_t^1 - V_t^1) - b_V c \epsilon_{t+1}^1 + k b_{\Theta} p_{\Theta} \epsilon_{t+1}^{\Theta})|\mathcal{F}_t^2] \\ &= \text{Var}[Q_{t+1}|\mathcal{F}_t^1] + \text{Var}[(ac(1-\lambda)p_{\Delta} - a(\lambda + (1-\lambda)p_{V^1}))(\hat{V}_t^1 - V_t^1)|\mathcal{F}_t^2] \\ &= \text{Var}[Q_{t+1}|\mathcal{F}_t^1] + a^2(\lambda + (1-\lambda)(p_{V^1} - c p_{\Delta}))^2 \text{Var}[\hat{V}_t^1 - V_t^1|\mathcal{F}_t^2]. \end{aligned}$$

The variance of the estimation error $\text{Var}[\hat{V}_t^1 - V_t^1|\mathcal{F}_t^2]$ can be found from the above theorem and solves the following equation:

$$\text{Var}[\hat{V}_t^1 - V_t^1|\mathcal{F}_t^2] = (1 - k(p_{V^1} - p_{\Delta}))(a^2 \text{Var}[\hat{V}_t^1 - V_t^1|\mathcal{F}_t^2] + b_V^2).$$

As a result,

$$\text{Var}[Q_{t+1}|\mathcal{F}_t^2] = \text{Var}[Q_{t+1}|\mathcal{F}_t^1] + (\lambda + (1-\lambda)(p_{V^1} - c p_{\Delta}))^2 \frac{a^2 c}{1 - a^2 c} b_V^2.$$

To summarize, we have the following system of nonlinear equations for p_{Θ} , p_{Δ} , c , ω_1 , ω_2 and Ω :

$$\begin{aligned} p_{\Theta} &= -\frac{1}{\Omega(1+r)}, \quad p_{\Delta} = \frac{\omega_2 \lambda a(1+r)}{(1+r-a(1-\lambda))(\Omega(1+r) - \omega_1 a c(1-\lambda))}, \\ \omega_1 (b_V^2 [(\lambda + (1-\lambda)(p_{V^1} - c p_{\Delta}))^2 + (\lambda + (1-\lambda)p_{V^2})^2] + b_{\Theta}^2(1-\lambda)^2 p_{\Theta}^2(1 + k p_{\Delta})^2) &= \gamma, \\ \omega_2 \left(b_V^2 \left[\frac{1}{1 - a^2 c} (\lambda + (1-\lambda)(p_{V^1} - c p_{\Delta}))^2 + (\lambda + (1-\lambda)p_{V^2})^2 \right] \right. \\ &\quad \left. + b_{\Theta}^2(1-\lambda)^2 p_{\Theta}^2(1 + k p_{\Delta})^2 \right) = 1 - \gamma, \\ p_{\Theta}^2 b_{\Theta}^2 a^2 (1-c)^2 + (p_{\Theta}^2 b_{\Theta}^2 (1-a^2) + b_V^2 (p_{V^1} - p_{\Delta})^2) (1-c) - b_V^2 (p_{V^1} - p_{\Delta})^2 &= 0, \end{aligned}$$

$$\Omega = \omega_1 + \omega_2.$$

We find the price function in terms of current values of state variables V_t^1 , V_t^2 , θ_t , and $(\hat{V}_t^1 - V_t^1)$. It also can be decomposed over shocks ϵ_{t-s}^1 , ϵ_{t-s}^2 , ϵ_{t-s}^Θ , $s \geq 0$. The result is as follows:

$$P_t = -\frac{1}{\Omega(r + \lambda)} + b_V \sum_{s=0}^{\infty} a^s (p_V - p_\Delta c^{s+1}) \epsilon_{t-s}^1 + b_V p_V \sum_{s=0}^{\infty} a^s \epsilon_{t-s}^2 \\ + b_\Theta p_\Theta (1 + p_\Delta k) \epsilon_t^\Theta + b_\Theta p_\Theta p_\Delta k \sum_{s=1}^{\infty} (ac)^s \epsilon_{t-s}^\Theta.$$

Appendix C

Proof of Proposition 3.

To save space we give the proof for $\alpha = 1$ and $\gamma = 1/2$ and the components V_t^1 and V_t^2 are treated symmetrically. The proof for the general case follows the same logic but is more involved. Denote a demeaned price as \tilde{P}_t . Assume that the model has a stationary equilibrium, i.e. \tilde{P}_t is a stationary regular Gaussian process¹⁹. It means that \tilde{P}_t admits the following decomposition:

$$\tilde{P}_t = b_V \sum_{k=0}^{\infty} f_k \epsilon_{t-k}^i + b_V \sum_{k=0}^{\infty} f_k \epsilon_{t-k}^{-i} + b_\Theta \sum_{k=0}^{\infty} f_k^\Theta \epsilon_{t-k}^\Theta,$$

where

$$\sum_{k=0}^{\infty} \left(b_V^2 f_k^2 + b_V^2 f_k^2 + b_\Theta^2 (f_k^\Theta)^2 \right) < \infty.$$

Instead of working with an infinite number of coefficients it is convenient to put the series in z -representation²⁰, i.e. introduce functions $f(z)$ and $f_\Theta(z)$ such that

$$f(z) = \sum_{k=0}^{\infty} f_k z^k, \quad f_\Theta(z) = \sum_{k=0}^{\infty} f_k^\Theta z^k.$$

Let L be a shift operator defined as $L\epsilon_t = \epsilon_{t-1}$. Then using z -representation we can put the conjectured price function into the following form:

$$\tilde{P}_t = b_V f(L) \epsilon_t^i + b_V f(L) \epsilon_t^{-i} + b_\Theta f^\Theta(L) \epsilon_t^\Theta.$$

Considering $f(z)$ and $f_\Theta(z)$ as functions in a complex plane we can conclude that they are analytic in the disk $D_0 = \{z : |z| < 1\}$. z -representation is convenient for taking

¹⁹See all relevant definitions in Ibragimov and Rozanov (1978).

²⁰For other applications of z -representation to analysis of rational expectation equilibrium see Futia (1981), Kasa (2000), Kasa, Walker, and Whiteman (2004).

expectations. If two random processes x_t and y_t are $x_t = b_V f_x^1(L)\epsilon_t^i + b_V f_x^2(L)\epsilon_t^{-i} + b_\Theta f_x^\Theta(L)\epsilon_t^\Theta$ and $y_t = b_V f_y^1(L)\epsilon_t^i + b_V f_y^2(L)\epsilon_t^{-i} + b_\Theta f_y^\Theta(L)\epsilon_t^\Theta$ then

$$E[x_t y_t] = b_V^2 \int_{|z|=1} \frac{1}{z} f_x^1(z) f_y^1\left(\frac{1}{z}\right) dz + b_V^2 \int_{|z|=1} \frac{1}{z} f_x^2(z) f_y^2\left(\frac{1}{z}\right) dz + b_V^2 \int_{|z|=1} \frac{1}{z} f_x^\Theta(z) f_y^\Theta\left(\frac{1}{z}\right) dz.$$

It turns out that the notion of Markovian dynamics has a nice counterpart in the frequency domain. We will use extensively the following result from the theory of Gaussian stationary processes (see Doob (1944) for original results and Ibragimov and Rozanov (1978) for textbook treatment).

Theorem. Let X_t be a regular Gaussian stationary process with discrete time defined on a complete probability space $(\Omega, \mathcal{F}, \mu)$. Let \mathcal{F}_t be a natural filtration generated by X_t . The process X_t admits Markovian dynamics with a finite number of Gaussian state variables if and only if its spectral density is a rational function $e^{i\lambda}$.

Remark. It is well-known result then that a Gaussian process X_t with a rational spectral density is an ARMA(p,q) process, that is, it can be represented as

$$X_t - \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (15)$$

for some $\phi_i, i = 1..p, \theta_i, i = 1..q$, and $\varepsilon_t, t \in \mathbb{Z}$.

It follows then (given (19)) that \tilde{P}_t will have a rational spectral density if and only if functions $f(z)$ and $f_\Theta(z)$ are rational functions as well.

Let us reformulate the equilibrium conditions in terms of functions $f(z)$ and $f_\Theta(z)$. It is convenient to start our consideration from the filtering problem of each agent. Indeed, while forming his demand each agent has to produce the best estimation of $\lambda V_{t+1}^{-i} + (1 - \lambda) P_{t+1}^i$ given his information set $\mathcal{F}_t^i = \{V_s^i, P_s\}_{-\infty}^t$. Since some components of P_t are known to agent i , observation of $\mathcal{F}_t^i = \{V_s^i, P_s\}_{-\infty}^t$ is equivalent to observation of $\mathcal{F}_t^i = \{V_s^i, Z_s^i\}_{-\infty}^t$, where

$$Z_t^i = b_V f(L)\epsilon_t^{-i} + b_\Theta f^\Theta(L)\epsilon_t^\Theta.$$

The filtering problem is equivalent to finding the projector G , such that:

$$E[\lambda V_{t+1}^{-i} + (1 - \lambda) Z_{t+1}^i | \mathcal{F}_t^i] = G(L) Z_t^i$$

By definition, $\lambda V_{t+1}^{-i} + (1 - \lambda) Z_{t+1}^i - G(L) Z_t^i$ is orthogonal to all $Z_s^i, s \leq t$:

$$E[(\lambda V_{t+1}^{-i} + (1 - \lambda) Z_{t+1}^i - G(L) Z_t^i) Z_s^i] = 0.$$

Calculating expectations we get

$$E[V_{t+1}^{-i} Z_s^i] = b_V^2 \int_{|z|=1} \frac{1}{z} \frac{a}{1 - az} \frac{1}{z^{t-s}} f\left(\frac{1}{z}\right) dz,$$

$$\begin{aligned}
E[Z_{t+1}^i Z_s^i] &= b_V^2 \int_{|z|=1} \frac{1}{z^2} f(z) \frac{1}{z^{t-s}} f\left(\frac{1}{z}\right) dz + b_\Theta^2 \int_{|z|=1} \frac{1}{z^2} f^\Theta(z) \frac{1}{z^{t-s}} f^\Theta\left(\frac{1}{z}\right) dz, \\
E[G(L)Z_t^i Z_s^i] &= \sum_{k=0}^{\infty} G_k E[Z_{t-k}^i Z_s^i] \\
&= \sum_{k=0}^{\infty} G_k \left(b_V^2 \int_{|z|=1} \frac{1}{z} f(z) \frac{1}{z^{t-k-s}} f\left(\frac{1}{z}\right) dz + b_\Theta^2 \int_{|z|=1} \frac{1}{z} f^\Theta(z) \frac{1}{z^{t-k-s}} f^\Theta\left(\frac{1}{z}\right) dz \right) \\
&= b_V^2 \int_{|z|=1} \frac{1}{z} \left(\sum_{k=0}^{\infty} G_k z^k \right) f(z) \frac{1}{z^{t-s}} f\left(\frac{1}{z}\right) dz + b_\Theta^2 \int_{|z|=1} \frac{1}{z} \left(\sum_{k=0}^{\infty} G_k z^k \right) f^\Theta(z) \frac{1}{z^{t-s}} f^\Theta\left(\frac{1}{z}\right) dz \\
&= b_V^2 \int_{|z|=1} \frac{1}{z} G(z) f(z) \frac{1}{z^{t-s}} f\left(\frac{1}{z}\right) dz + b_\Theta^2 \int_{|z|=1} \frac{1}{z} G(z) f^\Theta(z) \frac{1}{z^{t-s}} f^\Theta\left(\frac{1}{z}\right) dz.
\end{aligned}$$

Collecting all terms the orthogonality condition takes the form

$$\int_{|z|=1} \frac{1}{z^{t-s+1}} U(z) = 0$$

where the function $U(z)$ is

$$\begin{aligned}
U(z) &= b_V^2 \frac{a\lambda}{1-az} f\left(\frac{1}{z}\right) + (1-\lambda) \left(b_V^2 \frac{1}{z} f(z) f\left(\frac{1}{z}\right) + b_\Theta^2 \frac{1}{z} f_\Theta(z) f_\Theta\left(\frac{1}{z}\right) \right) - \\
&\quad - G(z) \left(b_V^2 f(z) f\left(\frac{1}{z}\right) + b_\Theta^2 f_\Theta(z) f_\Theta\left(\frac{1}{z}\right) \right).
\end{aligned}$$

It means that $U(z)$ is analytic in $D_\infty = \{z : |z| > 1\}$ and $U(\infty) = 0$. In other words, the series expansion of $U(z)$ at $z = \infty$ doesn't have the terms $z^s, s \geq 0$. In terms of projector G the demand function in z -representation is

$$\begin{aligned}
X_t^i &= -(r+\lambda)p_0 + b_V \left(\frac{a\lambda}{1-aL} - (1+r)f(L) + (1-\lambda) \frac{f(L) - f(0)}{L} \right) \epsilon_t^i \\
&\quad + b_V \left(-(1+r) + G(L) \right) f(L) \epsilon_t^{-i} + b_\Theta \left(-(1+r) + G(L) \right) f_\Theta(L) \epsilon_t^\Theta. \quad (16)
\end{aligned}$$

Market clearance condition $\omega_1 X_t^1 + \omega_2 X_t^2 = 1 + \theta_t$ where $\omega_1 = \omega_2 = \Omega/2$ should be valid for all realizations of shocks so it yields the following set of equations:

$$\begin{aligned}
\frac{a\lambda}{1-az} - 2(1+r)f(z) + (1-\lambda) \frac{f(z) - f(0)}{z} + G(z)f(z) &= 0, \\
-\Omega(1+r)f_\Theta(z) + \Omega G(z)f_\Theta(z) &= 1.
\end{aligned}$$

Given these equations $U(z)$ can be rewritten as

$$U(z) = 2b_V^2(1+r)f(z)f\left(\frac{1}{z}\right) + b_V^2(1-\lambda) \frac{f(0)}{z} f\left(\frac{1}{z}\right) - 2b_V^2 G(z)f(z)f\left(\frac{1}{z}\right) +$$

$$+b_{\Theta}^2(1-\lambda)\frac{1}{z}f_{\Theta}(z)f_{\Theta}\left(\frac{1}{z}\right)-b_{\Theta}^2\left(\frac{1}{2}+(1+r)f_{\Theta}(z)\right)f_{\Theta}\left(\frac{1}{z}\right).$$

Note that the term $b_V^2(1-\lambda)\frac{f(0)}{z}f\left(\frac{1}{z}\right)$ does not contribute the terms with non-negative powers of z , so it can be discarded. Similarly, the term $-\frac{1}{2}b_{\Theta}^2f_{\Theta}\left(\frac{1}{z}\right)$ contributes only the constant $-\frac{1}{2}b_{\Theta}^2f_{\Theta}(0)$. So $U(z)$ takes an equivalent form:

$$U(z) = 2b_V^2((1+r)-G(z))f(z)f\left(\frac{1}{z}\right) + b_{\Theta}^2\left((1-\lambda)\frac{1}{z}-(1+r)\right) \times \\ \times f_{\Theta}(z)f_{\Theta}\left(\frac{1}{z}\right) - \frac{1}{2}b_{\Theta}^2f_{\Theta}(0). \quad (17)$$

It is convenient to introduce $g(z) = G(z) - (1+r)$. Then the equations for $f(z)$, $f_{\Theta}(z)$ and $U(z)$ are

$$\frac{a\lambda}{1-az} - (1+r)f(z) + (1-\lambda)\frac{f(z)-f(0)}{z} + g(z)f(z) = 0, \quad f_{\Theta}(z) = \frac{1}{\Omega g(z)},$$

$$U = -2b_V^2g(z)f(z)f\left(\frac{1}{z}\right) + b_{\Theta}^2\left((1-\lambda)\frac{1}{z}-(1+r)\right)f_{\Theta}(z)f_{\Theta}\left(\frac{1}{z}\right) - \frac{1}{2}b_{\Theta}^2f_{\Theta}(0).$$

Equivalently,

$$f(z)\left((1-\lambda)\frac{1}{z}-(1+r)+g(z)\right) + \frac{a\lambda}{1-az} - (1-\lambda)\frac{f(0)}{z} = 0, \quad f_{\Theta}(z) = \frac{1}{\Omega g(z)},$$

$$U(z) = -2b_V^2g(z)f(z)f\left(\frac{1}{z}\right) + b_{\Theta}^2\left((1-\lambda)\frac{1}{z}-(1+r)+g(z)\right) \times \\ \times f_{\Theta}(z)f_{\Theta}\left(\frac{1}{z}\right) - b_{\Theta}^2f_{\Theta}(0). \quad (18)$$

or

$$f(z) = -\frac{a(\lambda+(1-\lambda)f(0))z-(1-\lambda)f(0)}{(1-az)(1-\lambda-(1+r)z+zg(z))}, \quad f_{\Theta}(z) = \frac{1}{\Omega g(z)}, \quad (19)$$

$$U(z) = -2b_V^2g(z)f(z)f\left(\frac{1}{z}\right) + b_{\Theta}^2(1-\lambda-(1+r)z+zg(z))\frac{1}{z} \times \\ \times f_{\Theta}(z)f_{\Theta}\left(\frac{1}{z}\right) - b_{\Theta}^2f_{\Theta}(0). \quad (20)$$

So the rational expectation equilibrium in our model is characterized by functions $f(z)$, $f_{\Theta}(z)$, $g(z)$ and $U(z)$ such that $f(z)$, $f_{\Theta}(z)$ and $g(z)$ are analytic inside the unit circle, $U(z)$ is analytic outside the unit circle and $U(\infty) = 0$ and equations (19) and (20) hold.

Now turn to the main part of the proof. The idea of our proof is based on Proposition 3. Indeed, it states that if the price process is Markovian then functions $f(z)$ and $f_{\Theta}(z)$ should

be rational. So to prove that the price process in our model is a non-Markovian we have to prove that there is no rational functions $f(z)$ and $f_{\Theta}(z)$ solving equations (19) and (20) and satisfying all conditions specified above. We construct the proof by contradiction. For further convenience we introduce the function $H(z)$ such that $g(z) = (zg(z) + 1 - \lambda - (1+r)z)H(z)$. Consequently, in terms of $H(z)$ the function $g(z)$ is

$$g(z) = (1+r) \frac{z_0 - z}{H^{-1}(z) - z} \quad \text{where } z_0 = \frac{1-\lambda}{1+r} \quad (21)$$

Given this we can specify the conditions on the function $H(z)$.

1. We know that $f_{\Theta}(z) = 1/(\Omega g(z))$. So

$$f_{\Theta}(z) = \frac{1}{\Omega(1+r)} \frac{1 - zH(z)}{(z_0 - z)H(z)}.$$

Since $f_{\Theta}(z)$ is rational and analytic in D_0 then $H(z)$ is rational and $H(z) \neq 0$ for $z \in D_0$.

Moreover,

$$H(z_0) = \frac{1}{z_0}.$$

2. We know that $g(z)$ is analytic D_0 . Consequently, the equation $1 - zH(z) = 0$ doesn't have other solutions in D_0 except $z = z_0$.

3. Substituting (21) into (19) gives

$$f(z) = \frac{a(\lambda + (1-\lambda)f(0))}{(1+r)} \frac{z_1 - z}{(1-az)(z_0 - z)} (1 - zH(z)) \quad \text{where } z_1 = \frac{(1-\lambda)f(0)}{a(\lambda + (1-\lambda)f(0))}.$$

Analyticity of $f(z)$ in D_0 implies that either $H(z)$ doesn't have poles in D_0 or the only pole is z_1 . In other words, $(z - z_1)H(z)$ is analytic in D_0 .

4. Substitution of $f(z)$ into $U(z)$ gives

$$U(z) = -2b_V^2 \frac{(a(\lambda + (1-\lambda)f(0))z - (1-\lambda)f(0))(a(\lambda + (1-\lambda)f(0))\frac{1}{z} - (1-\lambda)f(0))}{(1-az)(1-\frac{a}{z})g(\frac{1}{z})} \times \\ \times H(z)H\left(\frac{1}{z}\right) + b_{\Theta}^2 \frac{1}{\Omega^2 z H(z)g(\frac{1}{z})} - b_{\Theta}^2 f_{\Theta}(0). \quad (22)$$

Moreover,

$$f_{\Theta}(0) = \frac{1}{\Omega(1-\lambda)H(0)}$$

and since $g(z)$ doesn't have poles in D_0 (and consequently $g(\frac{1}{z})$ doesn't have poles in D_{∞}) analyticity of $U(z)$ in D_{∞} leads to analyticity of $U(z)g(\frac{1}{z})$ in D_{∞} . Abusing notations we get

$$U(z) = -2b_V^2 \frac{(a(\lambda + (1-\lambda)f(0))z - (1-\lambda)f(0))(a(\lambda + (1-\lambda)f(0))\frac{1}{z} - (1-\lambda)f(0))}{(1-az)(1-\frac{a}{z})} \times \\ \times H(z)H\left(\frac{1}{z}\right) + b_\Theta^2 \frac{1}{\Omega^2 z H(z)} - b_\Theta^2 \frac{1}{\Omega(1-\lambda)H(0)} g\left(\frac{1}{z}\right) \quad (23)$$

or

$$U(z) = -2b_V^2 a^2 (\lambda + (1-\lambda)f(0))^2 \frac{(z-z_1)(\frac{1}{z}-z_1)}{(1-az)(1-\frac{a}{z})} \times \\ \times H(z)H\left(\frac{1}{z}\right) + b_\Theta^2 \frac{1}{\Omega^2 z H(z)} - \frac{1}{\Omega} b_\Theta^2 \quad (24)$$

Analyticity of $U(z)$ in D_∞ means that the pole $1/a$ in (24) must be cancelled. It might happen only due to one of the following reasons:

- 1) $H(1/a) = 0$,
- 2) $H(a) = 0$,
- 3) $z_1 = a$, or, equivalently, $f(0) = \frac{a^2}{1-a^2} \frac{\lambda}{1-\lambda}$
- 4) $z_1 = 1/a$
- 5) pole in the first term is cancelled by pole in the second term.

It is easy to notice that the first reason doesn't work since in this case a pole in the second term appears. Similarly, the fifth possibility cannot realize. The equation $z_1 = 1/a$ is inconsistent until $\lambda = 0$. The second option contradicts the condition that $H(z)$ doesn't have zeros inside the unit circle. So the third possibility should realize and we can fix the value of $f(0)$. Consequently, we rewrite $U(z)$ as

$$U(z) = -2b_V^2 \frac{a^2 \lambda^2}{(1-a^2)^2} H(z)H\left(\frac{1}{z}\right) + b_\Theta^2 \frac{1}{\Omega^2 z H(z)} - \frac{1}{\Omega} b_\Theta^2$$

with the condition

$$H\left(\frac{1-\lambda}{1+r}\right) = \frac{1+r}{1-\lambda}.$$

Now we will show that there is no rational functions satisfying all specified conditions. Indeed, assume that $H(z)$ has a pole z_h . From condition 3 $z_h = a$ or $z_h \in D_\infty$. If $z_h \in D_\infty$ and $z_h \neq \infty$ then for analyticity of $U(z)$ in D_∞ we have to have $H(1/z_h) = 0$, but it contradicts to condition 1. If $z_h = a$ then $U(z)$ has a pole in $1/a$. Indeed, if a is a pole of $H(z)$ then $1/a$ is a pole of $H(1/z)$. The only possibility to cancel it in the first term of $U(z)$ is $H(1/a) = 0$. But in this case a pole in the second term arises. So $H(z)$ does not have poles in \mathbb{C} . So

the only possibility is $z_h = \infty$. Assume that $w_0 \in \mathbb{C}$ is a zero of $H(z)$: $H(w_0) = 0$. Due to condition 1 w_0 can be only in D_∞ . However, it means that until $H(1/z)$ or $H(z)$ have a pole at w_0 $U(z)$ is not analytic in D_∞ . However, we know that $H(z)$ (and consequently $H(1/z)$) do not have poles in \mathbb{C} . Thus we can conclude that $H(z)$ does not have zeros. The only possibility now is $H(z) = H = \text{const}$. We have two equations that this constant has to satisfy:

$$H = \frac{1+r}{1-\lambda}, \quad -2b_V^2 \frac{a^2 \lambda^2}{(1-a^2)^2} H^2 - \frac{1}{\Omega} b_\Theta^2 = 0.$$

Obviously, these conditions are inconsistent and this concludes the proof.

Appendix D

Given the complexity of the problem we use two numerical methods to solve the model. Both provide identical solutions.

k-lag revelation approximation

In the *k*-lag revelation approximation all information is revealed to all investors after *k* periods, so the information set of investor *i* is $\mathcal{F}_t^i = \{P_\tau, V_\tau^i : \tau \leq t; V_\tau^{-i}, \theta_\tau : \tau \leq t - k\}$. It means that the state of this economy Ψ_t is characterized by the current values of V_t^1, V_t^2 and θ_t and by their *k* lags:

$${}'\Psi_t = (\psi'_t, \psi'_{t-1}, \dots, \psi'_{t-k}, \psi'_{t-k})', \quad \text{where } \psi_\tau = (V_\tau^1, V_\tau^2, \theta_\tau)'.$$

Demand of type *i* investors is

$$X_t^i = \omega_i E[Q_{t+1} | \mathcal{F}_t^i] = \omega_i (a\lambda V_t^i - (1+r)P_t + E[a\lambda V_t^{-i} + (1-\lambda)P_{t+1} | \mathcal{F}_t^i]),$$

where ω_i are endogenous constants given by

$$\omega_1 = \frac{\gamma}{\alpha \text{Var}[Q_{t+1} | \mathcal{F}_t^1]}, \quad \omega_2 = \frac{1-\gamma}{\alpha \text{Var}[Q_{t+1} | \mathcal{F}_t^2]}$$

We look for the equilibrium price process as a linear function of state variables, i.e. $P_t = p_0 + P\Psi_t$, where P is a $(1 \times 3(k+1))$ constant matrix. In the matrix form dynamics of ψ_t is:

$$\psi_{t+1} = a_\psi \psi_t + \epsilon_{t+1}^\psi, \quad \text{where } a_\psi = \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \epsilon_t^\psi = \begin{pmatrix} \epsilon_t^1 \\ \epsilon_t^2 \\ \epsilon_t^\Theta \end{pmatrix}, \quad \text{Var}(\epsilon_t^\psi) = \begin{pmatrix} b_V^2 & 0 & 0 \\ 0 & b_V^2 & 0 \\ 0 & 0 & b_\Theta^2 \end{pmatrix}.$$

Consequently, dynamics of Ψ_t can be described as:

$$\Psi_{t+1} = A_\Psi \Psi_t + B_\Psi \epsilon_{t+1}^\psi, \quad \text{where} \quad A_\Psi = \begin{pmatrix} a_\psi & 0 & \dots & 0 & 0 \\ I_3 & 0 & \dots & 0 & 0 \\ 0 & I_3 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_3 & 0 \end{pmatrix}, \quad B_\Psi = \begin{pmatrix} I_3 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Here I_3 is a 3-dimensional unit matrix. Now demand can be rewritten as

$$X_t^i = \omega_i((1 - \lambda)p_0 + a\lambda V_t^i - (1 + r)P_t + E[a\lambda V_t^{-i} + (1 - \lambda)PA_\Psi \Psi_t | \mathcal{F}_t^i]).$$

Introducing $(1 \times 3(k + 1))$ constant matrices $V^1 = (1, 0, 0, \dots, 0)$, $V^2 = (0, 1, 0, \dots, 0)$ and $V = (1, 1, 0, \dots, 0)$ we get

$$X_t^i = \omega_i((1 - \lambda)p_0 - (1 + r)P_t) + \omega_i(a\lambda V + (1 - \lambda)PA_\Psi)E[\Psi_t | \mathcal{F}_t^i].$$

Thus, we have to calculate $E[\Psi_t | \mathcal{F}_t^i]$. Denoting time t observations of agent i as $y_t^i = (P_t, V_t^i)'$ we can gather all his relevant observations in one vector $Y_t^i = (y_t^{i'}, y_{t-1}^{i'}, \dots, y_{t-k+1}^{i'}, \psi_{t-k})$. It is also convenient to introduce a set of \tilde{P}_τ , $\tau = t - k + 1 \dots t$ to separate the informative part of the price:

$$\begin{aligned} \tilde{P}_t &= P_t - p_0, \\ \tilde{P}_{t-1} &= P_{t-1} - p_0 - P^k \psi_{t-k-1}, \\ &\dots \\ \tilde{P}_{t-k+1} &= P_{t-k+1} - p_0 - P^2 \psi_{t-k-1} - \dots - P^k \psi_{t-2k}. \end{aligned}$$

Now we can put all observations in an equivalent matrix form:

$$Y_t^i = H^i \Psi_t, \quad \text{where} \quad H^i = \begin{pmatrix} h^i \\ h^i J \\ h^i J^2 \\ \vdots \\ h^i J^k \\ O_{3 \times 3k} \quad I_3 \end{pmatrix}, \quad J = \begin{pmatrix} 0 & I_3 & 0 & \dots & 0 \\ 0 & 0 & I_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & I_3 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad h^i = \begin{pmatrix} P \\ V^i \end{pmatrix}$$

To infer $E[\Psi_t | \mathcal{F}_t^i]$ we use the following well-known fact: if random variables (Ψ, Y) are jointly normal with zero mean, then

$$E[\Psi | Y] = \beta' Y, \quad \text{where} \quad \beta = \text{Var}(Y)^{-1} E(Y \Psi'), \quad (25)$$

$$\text{Var}[\Psi|Y] = \text{Var}(\Psi) - E(Y\Psi')'\text{Var}(Y)^{-1}E(Y\Psi'). \quad (26)$$

In our particular case we have:

$$\text{Var}(Y_t^i) = H^i \text{Var}(\Psi_t) H^{i'}$$

$$E(Y^i \Psi') = H^i \text{Var}(\Psi_t)$$

From the dynamic equation for Ψ_t we find that

$$\text{Var}(\Psi_t) = A_\Psi \text{Var}(\Psi_t) A_\Psi' + B_\Psi \text{Var}(\epsilon_t^\psi) B_\Psi'$$

Iterating we get

$$\text{Var}(\Psi_t) = \sum_{l=0}^{\infty} A_\Psi^l B_\Psi \text{Var}(\epsilon_t^\psi) B_\Psi' A_\Psi^{l'}$$

Thus, the demand of agent i is

$$X_t^i = \omega_i((1-\lambda)p_0 - (1+r)P_t) + \omega_i(a\lambda V + (1-\lambda)PA_\Psi) \text{Var}(\Psi_t) H^{i'} (H^i \text{Var}(\Psi_t) H^{i'})^{-1} H^i \Psi_t.$$

The market clearing condition gives

$$\begin{aligned} & \Omega(1-\lambda)p_0 - \Omega(1+r)P_t \\ & + \omega_1(a\lambda V + (1-\lambda)PA_\Psi) \text{Var}(\Psi_t) H^{1'} (H^1 \text{Var}(\Psi_t) H^{1'})^{-1} H^1 \Psi_t \\ & + \omega_2(a\lambda V + (1-\lambda)PA_\Psi) \text{Var}(\Psi_t) H^{2'} (H^2 \text{Var}(\Psi_t) H^{2'})^{-1} H^2 \Psi_t = 1 + \Theta \Psi_t, \end{aligned} \quad (27)$$

where $\Theta = (0, 0, 1, 0, 0, \dots, 0)$ and $\Omega = \omega_1 + \omega_2$. Rearranging terms we get:

$$\begin{aligned} P_t &= \frac{\Omega(1-\lambda)p_0 - 1}{\Omega(1+r)} + \frac{1}{\Omega(1+r)} (a\lambda V + (1-\lambda)PA_\Psi) \text{Var}(\Psi_t) \times \\ & \times \left[\omega_1 H^{1'} (H^1 \text{Var}(\Psi_t) H^{1'})^{-1} H^1 + \omega_2 H^{2'} (H^2 \text{Var}(\Psi_t) H^{2'})^{-1} H^2 \right] \Psi_t - \frac{1}{\Omega(1+r)} \Theta \Psi_t. \end{aligned} \quad (28)$$

Comparing this equation with the price representation $P_t = p_0 + P\Psi_t$ we get a set of equations:

$$p_0 = \frac{\Omega(1-\lambda)p_0 - 1}{\Omega(1+r)}, \quad \text{or} \quad p_0 = -\frac{1}{\Omega(r+\lambda)},$$

$$\begin{aligned} \Omega(1+r)P &= (a\lambda V + (1-\lambda)PA_\Psi) \text{Var}(\Psi_t) \times \\ & \times \left(\omega_1 H^{1'} (H^1 \text{Var}(\Psi_t) H^{1'})^{-1} H^1 + \omega_2 H^{2'} (H^2 \text{Var}(\Psi_t) H^{2'})^{-1} H^2 \right) - \Theta. \end{aligned} \quad (29)$$

This system of equations on matrix P should be supplemented by two equations determining ω_1 and ω_2 . By definition, ω_i are determined by conditional variances $\text{Var}[Q_{t+1}|\mathcal{F}_t^i]$

$$\text{Var}[Q_{t+1}|\mathcal{F}_t^i] = \text{Var}[\lambda V_{t+1} + (1-\lambda)P_{t+1}|\mathcal{F}_t^i] = \text{Var}[(\lambda V + (1-\lambda)P)\Psi_{t+1}|\mathcal{F}_t^i]$$

$$= (\lambda V + (1 - \lambda)P)\text{Var}[\Psi_{t+1}|\mathcal{F}_t^i](\lambda V + (1 - \lambda)P)'$$

From (26) we get

$$\text{Var}[\Psi_{t+1}|\mathcal{F}_t^i] = A_\Psi(\text{Var}(\Psi_t) - \text{Var}(\Psi_t)H^{i'}(H^i\text{Var}(\Psi_t)H^{i'})^{-1}H^{i'}\text{Var}(\Psi_t))A'_\Psi + B_\Psi\text{Var}(\epsilon_t^\psi)B'_\Psi.$$

So the additional equations are

$$\begin{aligned} \frac{1}{\omega_i} &= (\lambda V + (1 - \lambda)P)(A_\Psi(\text{Var}(\Psi_t) - \text{Var}(\Psi_t)H^{i'}(H^i\text{Var}(\Psi_t)H^{i'})^{-1}H^{i'}\text{Var}(\Psi_t))A'_\Psi \\ &\quad + B_\Psi\text{Var}(\epsilon_t^\psi)B'_\Psi)(\lambda V + (1 - \lambda)P)'. \end{aligned} \quad (30)$$

As a result, when all information is revealed after k lags the equilibrium condition transforms into a complicated system of non-linear equations(29) and (30) determining P , ω_1 and ω_2 . Numerical solution to these equations give us an approximation to the original heterogeneous information equilibrium.

k -lag approximation allows us to calculate explicitly the decomposition of higher order expectations over the state variables Ψ_t . Indeed,

$$\begin{aligned} \bar{E}_t^w[V_t] &= \frac{1}{\Omega} (\omega_1 E[V_t|\mathcal{F}_t^1] + \omega_2 E[V_t|\mathcal{F}_t^2]) \\ &= \frac{1}{\Omega} (\omega_1 V E[\Psi_t|\mathcal{F}_t^1] + \omega_2 V E[\Psi_t|\mathcal{F}_t^2]) = \frac{1}{\Omega} V (\omega_1 \Pi^1 + \omega_2 \Pi^2) \Psi_t, \end{aligned} \quad (31)$$

where

$$\Pi^i = \text{Var}(\Psi_t)H^{i'} \left(H^i \text{Var}(\Psi_t) H^{i'} \right)^{-1} H^i.$$

Iterating we get:

$$\begin{aligned} \bar{E}_t^w \bar{E}_{t+1}^w[V_{t+1}] &= \bar{E}_t^w \left[\frac{1}{\Omega} V (\omega_1 \Pi^1 + \omega_2 \Pi^2) \Psi_{t+1} \right] \\ &= \frac{1}{\Omega} V (\omega_1 \Pi^1 + \omega_2 \Pi^2) A_\Psi \frac{1}{\Omega} (\omega_1 \Pi^1 + \omega_2 \Pi^2) \Psi_t, \end{aligned}$$

...

$$\bar{E}_t^w \bar{E}_{t+1}^w \dots \bar{E}_{t+s}^w[V_{t+s}] = \frac{1}{\Omega} V (\omega_1 \Pi^1 + \omega_2 \Pi^2) \left[A_\Psi \frac{1}{\Omega} (\omega_1 \Pi^1 + \omega_2 \Pi^2) \right]^s \Psi_t$$

Projection approximation

Projection approximation is another finite dimensional approximation to an infinite dimensional dynamic system describing the equilibrium in our model. Consider all random variables that admit the following decomposition:

$$x = b_V \sum_{k=0}^{\infty} f_k^1 \epsilon_{t-k}^1 + b_V \sum_{k=0}^{\infty} f_k^2 \epsilon_{t-k}^2 + b_{\Theta} \sum_{k=0}^{\infty} f_k^{\Theta} \epsilon_{t-k}^{\Theta},$$

where

$$\sum_{k=0}^{\infty} \left(b_V^2 (f_k^1)^2 + b_V^2 (f_k^2)^2 + b_{\Theta}^2 (f_k^{\Theta})^2 \right) < \infty$$

and denote the set of such random variables as H . From Appendix C we know that the demeaned price process \tilde{P} is in this set: $\tilde{P} \in H$. In fact, H is a Hilbert space with a scalar product defined as follows. If $x \in H$ and $y \in H$ can be decomposed over shocks with the coefficients $(f_k^1, f_k^2, f_k^{\Theta})_{k=0}^{\infty}$ and $(g_k^1, g_k^2, g_k^{\Theta})_{k=0}^{\infty}$ respectively, then the scalar product is

$$(x, y) = E[xy] = \sum_{k=0}^{\infty} (b_V^2 f_k^1 g_k^1 + b_V^2 f_k^2 g_k^2 + b_{\Theta}^2 f_k^{\Theta} g_k^{\Theta}).$$

In what follows we describe each element $x \in H$ by its coefficients $(f_k^1, f_k^2, f_k^{\Theta})_{k=0}^{\infty}$. Note that in this representation V^1, V^2 , and θ are $(b_V, b_V a, b_V a^2, \dots; 0, 0)$, $(0; b_V, b_V a, b_V a^2, \dots; 0)$, and $(0; 0; b_{\Theta})$ respectively. It is convenient to introduce a shift operator L such as

$$L(f_k^1, f_k^2, f_k^{\Theta})_{k=0}^{\infty} = (\hat{f}_k^1, \hat{f}_k^2, \hat{f}_k^{\Theta})_{k=0}^{\infty},$$

where

$$\hat{f}_0^1 = 0, \hat{f}_0^2 = 0, \hat{f}_0^{\Theta} = 0, \text{ and } \hat{f}_{k+1}^1 = f_k^1, \hat{f}_{k+1}^2 = f_k^2, \hat{f}_{k+1}^{\Theta} = f_k^{\Theta} \text{ for } k = 0, 1, \dots$$

Using this operator we can represent the demeaned excess return \tilde{Q} as $\tilde{Q} = \lambda V + (1 - \lambda)\tilde{P} - (1 + r)L\tilde{P}$. Let $M \subseteq H$ be a linear subspace and define $\pi_{\{M\}}$ as a projection operator on M . Denote the projection operator of each agent on his information set as π^i : $\pi^i = \pi_{\{LV^i, L^2V^i, \dots, L\tilde{P}, L^2\tilde{P}, \dots\}}$. Then we have $\omega_i = \frac{1}{\alpha \|\tilde{Q} - \pi^i \tilde{Q}\|^2}$ and the equilibrium price \tilde{P} is such that the following equation is satisfied:

$$(\gamma \omega_1 \pi^1 + (1 - \gamma) \omega_2 \pi^2) \tilde{Q} = \theta. \quad (32)$$

While this equilibrium condition might appear simple it should be understood as an infinite dimensional system of non-linear equations in the Hilbert space H which determines the coefficients $(p_k^1, p_k^2, p_k^{\Theta})_{k=0}^{\infty}$ of the price process \tilde{P} . To tackle this problem we consider a

sequence of finite dimensional approximations. Instead of an infinite dimensional system (32) we consider a finite dimensional one:

$$(\gamma\omega_1^N\pi_N^1 + (1-\gamma)\omega_2^N\pi_N^2)\tilde{Q} = \theta. \quad (33)$$

Here $\pi_N^i = \pi_{\{\pi_N L V^i, \pi_N L^2 V^i, \dots, \pi_N L^N V^i, \pi_N L \tilde{P}, \pi_N L^2 \tilde{P}, \dots, \pi_N L^N \tilde{P}\}}$ is a finite dimensional analog of π^i and it projects the elements of H on a finite dimensional subspace generated by vectors $\{\pi_N L V^i, \pi_N L^2 V^i, \dots, \pi_N L^N V^i, \pi_N L \tilde{P}, \pi_N L^2 \tilde{P}, \dots, \pi_N L^N \tilde{P}\}$. π_N is a projection operator on the space spanned by the first N basis vectors, which correspond to N most recent shocks in the initial space: $\pi_N(f_k^1, f_k^2, f_k^\Theta)_{k=0}^\infty = (f_k^1, f_k^2, f_k^\Theta)_{k=0}^N$. Correspondingly, $\omega_i^N = 1/(\alpha\|\tilde{Q} - \pi_N^i \tilde{Q}\|^2)$. Thus, instead of an infinite dimensional problem (32) we solve (33). We demonstrate that as N tends to infinity the approximations are more and more closer to each other. This fact indicates that as $N \rightarrow \infty$ the approximate solution converges to the real one.

References

- [1] Admati A. R. (1985) A Noisy Rational Expectations Equilibrium for Multi-asset Securities Markets, *Econometrica* 53, 629-657
- [2] Admati A. R. and P. Pfleiderer (1986) A monopolistic market for information, *Journal of Economic Theory* 39, 400-438
- [3] Allen F., S. Morris and H. S. Shin (2004) Beauty Contests, Bubbles and Iterated Expectations in Asset Markets, Cowles Foundation Discussion Paper No. 1406
- [4] Bacchetta, Ph., and E. van Wincoop (2003) Can Information Heterogeneity Explain the Exchange Rate Determination Puzzle? NBER WP. 9498
- [5] Bacchetta, Ph., and E. van Wincoop (2004) Higher Order Expectations in Asset Pricing, mimeo
- [6] Back K., C. H. Cao and G. A. Willard (2000) Imperfect Competition among Informed Traders, *Journal of Finance* 55, 2117-2155
- [7] Barberis N., A. Shleifer and R. Vishny (1998) A Model of Investor Sentiment, *Journal of Financial Economics* 49, 307-343
- [8] Berk J. B., R. C. Green, and V. Naik (1999) Optimal Investment, Growth Options, and Security Returns, *Journal of Finance* 54, 1553-1607
- [9] Bernhardt, D., P. Seiler, and B. Taub (2004) Speculative dynamics, working paper, University of Illinois
- [10] Brown, D. P. and R. H. Jennings (1989) On Technical Analysis, *Review of Financial Studies* 2, 527-552
- [11] Brunnermeier, M. (2001) *Asset Pricing Under Asymmetric Information*, Oxford University Press
- [12] Campbell, J. Y. and S. J. Grossman, and J. Wang (1993) Trading Volume and Serial Correlation in Stock Returns, *Quarterly Journal Economics* 108, 905-939
- [13] Campbell, J. Y., and A. S. Kyle (1993) Smart Money, Noise Trading, and Stock Price Behavior, *Review of Economic Studies* 60, 1-34

- [14] Daniel K., D. Hirshleifer and A. Subrahmanyam (1998) Investor Psychology and Security Market under-and Overreaction, *Journal of Finance* 53, 1839-1885
- [15] Doob J. L. (1944) The elementary Gaussian processes, *Annals of Mathematical Statistics* 15, 229-282
- [16] Easley, D., and M. O'Hara (2004) Information and the cost of capital, *Journal of Finance* 59, 1553-1583
- [17] Foster F. D. and S. Viswanathan (1996) Strategic Trading When Agents Forecast the Forecasts of Others, *Journal of Finance* 51, 1437-1477
- [18] Futia, C. A. (1981) Rational Expectations in Stationary Linear Models, *Econometrica* 49, 171-192
- [19] Grossman, S. J. and J. E. Stiglitz (1980) On the Impossibility of Informationally Efficient Markets, *American Economic Review* 70, 393-408
- [20] Grundy, B. D. and M. McNichols (1989) Trade and Revelation of Information through Prices and Direct Disclosure, *Review of Financial Studies* 2, 495-526
- [21] Grundy, B. D. and Y. Kim (2002) Stock Market Volatility in a Heterogenous Information Economy, *Journal of Financial and Quantitative Analysis* 37, 1-27
- [22] He, H., and J. Wang (1995) Differential Information and Dynamic Behavior of Stock Trading Volume, *The Review of Financial Studies* 8, 919-972
- [23] Hong, H. and J. Stein (1999) A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets, *Journal of Finance* 54, 2143-2184
- [24] Hong, H., T. Lim, and J. Stein (2000) Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies, *Journal of Finance* 55, 265-295
- [25] Hughes, J., J. Liu, and J. Liu (2005) Information, Diversication, and Cost of capital, Working paper
- [26] Ibragimov I.A., and Y.A. Rozanov (1978) *Gaussian Random Processes*, Springer-Verlag
- [27] Jazwinski, A. H. (1970) *Stochastic Processes and Filtering Theory*, Academic Press

- [28] Jegadeesh N. and S. Titman (1993) Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance* 48, 65-91
- [29] Jegadeesh N. and S. Titman (2002) Cross-Sectional and Time-Series Determinants of Momentum Returns *Review of Financial Studies* 15, 143 - 157.
- [30] Jegadeesh N. and S. Titman (2005) Momentum, in R. H. Thaler (editor), *Advances in Behavioral Finance*, Volume 2, Princeton University Press
- [31] Johnson, T. C. (2002) Rational Momentum Effects, *Journal of Finance* 57, 585-607
- [32] Kasa, K. (2000) Forecasting the Forecasts of Others in the Frequency Domain, *Review of Economic Dynamics* 3, 726 - 756
- [33] Kasa, K., T. B. Walker and C. H. Whiteman (2004) Asset Pricing with Heterogeneous Beliefs: A Frequency Domain Approach, mimeo
- [34] Keynes, J. M. (1936) *The General Theory of Employment, Interest and Money*, MacMillan, London
- [35] Lo, A. W. and J. Wang (2000) Trading Volume: Definitions, Data Analysis, and Implications of Portfolio Theory, *Review of Financial Studies* 13, 257-300
- [36] Lo, A. and A. C. MacKinlay (1990) When are Contrarian Profits Due to Stock Market Overreaction?, *Review of Financial Studies* 3, 175-208.
- [37] Lee, C. M. C. and B. Swaminathan (2000) Price Momentum and Trading Volume, *Journal of Finance* 55, 2017-2069
- [38] Naik, N. Y. (1997a) On aggregation of information in competitive markets: The dynamic case, *Journal of Economic Dynamics and Control* 21, 1199-1227
- [39] Naik, N. Y. (1997b) Multi-period information markets, *Journal of Economic Dynamics and Control* 21, 1229-1258
- [40] Pearlman, J. G. and T. J. Sargent (2004) Knowing the Forecasts of Others, mimeo, New York University
- [41] Sargent, T. J. (1991) Equilibrium with Signal Extraction from Endogenous Variables, *Journal of Economic Dynamics and Control* 15, 245-273

- [42] Schroder M. and C. Skiadas (1999) Optimal Consumption and Portfolio Selection with Stochastic Differential Utility, *Journal of Economic Theory* 89, 68-126
- [43] Singleton K. J. (1987) Asset prices in a time-series model with disparately informed, competitive traders, in Barnett, W. A. and Singleton , K. J. (eds.), *New Approaches to Monetary Economics, Proceedings of the Second International Symposium in Economic Theory and Econometrics*, Cambridge University Press
- [44] Townsend, R. M. (1983a) Forecasting the Forecasts of Others, *Journal of Political Economy* 91, 546 - 588
- [45] Townsend, R. M. (1983b) Equilibrium theory with learning and disparate expectations: some issues and methods, in R. Frydman and E. Phelps (eds), *Individual Forecasting and Aggregate Outcomes*, Cambridge University Press
- [46] Verrecchia, R. E. (1982) Information Acquisition in a Noisy Rational Expectations Economy, *Econometrica* 50, 1415-1430
- [47] Verardo, M. (2005) Heterogeneous Beliefs and Momentum Profits, Working paper
- [48] Vives, X. (1988) Aggregation of Information in Large Cournot Markets, *Econometrica* 56, 851-876.
- [49] Wang, J. (1993) A Model of Intertemporal Asset Prices Under Asymmetric Information, *Review of Economic Studies* 60, 249-282
- [50] Wang, J. (1994) A Model of Competitive Stock Trading Volume, *Journal of Political Economy* 102, 127-168
- [51] Woodford, M. (2002) Imperfect Common Knowledge and the Effects of Monetary Policy, in P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford (eds.), *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, Princeton Univ. Press.