

# Smooth Trading with Overconfidence and Market Power

ALBERT S. KYLE, ANNA A. OBIZHAEVA, AND YAJUN WANG\*

First Draft: June 14, 2012

This Draft: October 21, 2013

*We describe a symmetric continuous-time model of trading among relatively overconfident oligopolistic informed traders with exponential utility. Traders agree to disagree about the precisions of their continuous flows of Gaussian private information. With enough disagreement, an equilibrium exists in which prices reveal the average of all traders' signals immediately, but traders continue to trade gradually towards target inventories. The price is a linear function of a trader's inventory, the derivative of a trader's inventory, and the average other traders' valuations. Prices reflect a "Keynesian beauty contest." Faster-than-equilibrium trading generates "flash crashes."*

*JEL:* D8, D43, D47, G02, G14

*Keywords:* market microstructure, price impact, liquidity, transactions costs, strategic trading, dynamic trading, double auctions, information aggregation, rational expectations, agreement to disagree, imperfect competition, bid shading, overconfidence, econophysics, asset management, flash crash, Keynesian beauty contest

\* Kyle: Robert H. Smith School of Business, University of Maryland, College Park, MD 20742, akyle@rsmith.umd.edu. Obizhaeva: Robert H. Smith School of Business, University of Maryland, College Park, MD 20742, obizhaeva@rsmith.umd.edu. Wang: Robert H. Smith School of Business, University of Maryland, College Park, MD 20742, ywang22@rsmith.umd.edu. We thank Vincent Fardeau, Jungsuk Han, Jeongmin Lee, Wei Li, Hong Liu, Mark Loewenstein, Ioanid Rosu, Dimitri Vayanos, Bin Wei, and Liyan Yang for their comments.

When large traders in financial markets seek to profit from perishable private information while keeping transactions cost low, they face a fundamental trade-off. On the one hand, they want to trade slowly, to reduce their own temporary price impact costs resulting from adverse selection. On the other hand, they want to trade quickly, before the permanent price impact of competitors trading on similar information makes temporary profit opportunities go away. We describe a model of continuous information-based trading which illustrates precisely this tradeoff between the temporary price impact costs of a trader's own trades and the permanent price impact of other traders acting on similar information.

The model has the following key features: (1) There is only one type of trader—a strategic informed trader; there are no noise traders or market makers. (2) Each trader has private information about the same underlying fundamental value; the “noise” in their signals is uncorrelated. (3) Traders are “relatively overconfident,” in the sense that each trader believes his private information is more precise than other traders believe it to be. (4) Given his beliefs about the precision of his and others’ signals, each trader applies Bayes Law correctly; in doing so, each trader infers from prices the economically relevant aggregation of all other traders’ information. (5) Traders trade strategically, correctly taking into account how the permanent and temporary price impact of their trades affects the trading of other traders. (6) Random variables are jointly normally distributed and traders have additive exponential utility functions. (7) The traders are “symmetric” in the sense that they have the same utility functions and symmetrically different beliefs about the information structure in the economy.

We describe an “almost closed form” steady state equilibrium with “smooth trading”<sup>1</sup> characterized precisely by endogenous parameters solving a set of six polynomial equations in six unknowns. In this equilibrium, each trader correctly believes that the market price is a linear function of (1) an average of other traders’ buy-and-hold valuations, (2) the trader’s own inventory, and (3) the derivative of the trader’s own inventory. The one-period version of the model has a closed-form solution implying existence of an equilibrium with positive trading volume when each trader believes the square root of the precision of his signal is slightly greater than twice that of other traders. Numerical calculations indicate that the same existence condition holds in the continuous-time model.

The model explains the following realistic features of trading in speculative markets.

First, since the market offers no instantaneous liquidity for block trades, traders must “shred orders” so that inventories are differentiable functions of time. This is consistent with the real world fact that equity asset managers limit trading to, say, five percent of daily volume, perhaps using trades based on volume-weighted-average-prices (VWAP) to acquire positions of days, weeks, or even months.

<sup>1</sup>Our informal use of the term “smooth trading” is different from the mathematical sense that implies derivatives of all orders exist. Since the first derivatives of traders’ inventories follow diffusions, higher order derivatives do not exist.

Second, while prices reveal new information immediately, trading on the basis of innovations in signals continues after the information is revealed in prices. Each trader's inventory follows a partial adjustment model which adjusts positions linearly in the direction of a "target inventory."

Third, because strategic traders compete with one another to acquire private information about the same underlying fundamental value, they build and reduce positions over endogenously determined horizons which are much shorter than the horizons over which the long-term fundamentals unfold. For example, an equity asset manager who obtains information about the long-term ten-year growth rate of a firm's cash flows might build up a position gradually over several weeks, then liquidate the position several weeks or months later, after his information becomes generally known to others.

Fourth, trading exhibits precisely the features of the "beauty contest" described by Keynes (1936), with the exception that the beauty contest dampens price fluctuations rather than exacerbates them. Prices are a weighted average of the buy-and-hold valuations of each trader; "dampening" implies that the weights sum to a constant less than one. Intuitively, the conditions under which the beauty contest would exacerbate price fluctuations are precisely the conditions under which market equilibrium fails to exist.

Fifth, abnormally fast or "out-of-equilibrium" execution of bets, such as fire sales, can result in price spikes resembling a "flash crash." Prices collapse before positions are actually sold and recover while the selling takes place. This occurs because prices react immediately to changes in the derivatives of a trader's inventory.

Sixth, the valuation of the risky security in each trader's utility function adjusts the market price for the asset's illiquidity and for the trader's private estimate of its value.

Finally, the steady-state level of market liquidity reflects a delicate and unstable balance between the eagerness of traders to demand it from the market and their willingness to supply it to the market.

### **Related Literature.**

The one-period version of our model implements an equilibrium in demand curves, like Kyle (1989) and Rostek and Weretka (2012). Whether based on our agreement-to-disagree assumption or the private values assumption of Rostek and Weretka (2012), the bid shading which results from strategic bidding implies that equilibrium only exists if there is enough disagreement or a sufficient difference in private values to support it. The continuous-time model implements a continuous version of the auction mechanism in Kyle (1989) in which traders continuously submit demand schedules. The schedules are for flows—derivatives of inventories which define the speed of trading—rather than for levels of inventories.

The concepts of depth, tightness, and resiliency from Black (1971) play out differently in our model than in Kyle (1985). The continuous-time model of Kyle (1985) works as follows. Market makers earn spread profits from infinitely impatient noise

traders whose inventories follow diffusions and are therefore not differentiable. This allows market makers to provide market depth to all traders in such a manner that trades have permanent price impact but no transitory price impact. The equilibrium would be broken if the noise traders were to cut bid-ask spread costs in half by smoothing their trading over very short periods of time, walking the demand curve of the market makers. By assumption, they do not do so. The informed trader's private information does not decay. His trades lead to permanent price impact but no temporary price impact. He smooths his trading, but the evenness with which his trading is smoothed is not derived from his profit maximization problem; instead, it is derived from the need to sustain an equilibrium with constant market depth. Constant market depth implies that permanent price impact does not depend on how the informed trader spreads his trades out over time.

In our equilibrium, in contrast with Kyle (1985), all traders smooth their trading so that the derivatives of inventories are differentiable; thus, no traders provide instantaneous liquidity to large blocks. By definition, this makes trading volume finite, in contrast to the infinite trading volume resulting from inventories of noise traders and market makers following diffusions as in Kyle (1985). More importantly, smooth trading changes the manner in which liquidity is provided in a profound manner. From the perspective of each trader, the price has a permanent price impact term linear in the level of the trader's inventory and a transitory price impact term linear in the derivative of the trader's inventory. The permanent price impact term is analogous to the price impact quantified by the "lambda" parameter in Kyle (1985). Because of the transitory price impact term, trading a given quantity over a shorter horizon is more expensive than trading it over a longer horizon. This gives informed traders an incentive to slow down their trading. Since their private information decays, they have an offsetting incentive to speed up their trading. The equilibrium balances this trade-off.

Our continuous-time model is closest to Vayanos (1999) and Du and Zhu (2013), who both present dynamic models in which symmetric strategic oligopolistic traders smooth out their trading to reduce market impact resulting from adverse selection. Their models are set in discrete time; our continuous-time approach brings out clearly the intuition that traders trade smoothly. The equilibrium implications of our model are significantly different from the models of both Vayanos (1999) and Du and Zhu (2013) due to differences concerning the orthogonality of private information. Vayanos (1999) assumes that traders receive orthogonal endowment shocks and trade for risk-sharing motives; the orthogonal endowment shocks are a form of private information. Du and Zhu (2013) assume that traders' valuations have orthogonal "private values" components; these orthogonal private values components are also a form of private information. In our model, traders receive noisy signals about the same underlying fundamental; this makes their signals have positive correlation. The economically important implications of this positive correlation are precisely what our model is designed to capture. Because of the intrinsic strategic interactions resulting from positive correlation in signals in a context where traders agree

to disagree, our equilibrium is more difficult to characterize and has substantially different properties. Since traders agree to disagree about the information structure, each trader not only applies Bayes Law differently but also takes into account the different ways in which other traders apply Bayes Law. In Vayanos (1999) and Du and Zhu (2013), the horizon over which traders smooth buying and selling depends on their risk aversion, not the correlation of their private signals. Their result is consistent with Grossman and Miller (1988), who relate traders' impatience to risk aversion. In our model, the horizon over which traders smooth buying and selling depends strongly on the rate at which their signals decay due to information acquisition by others. Since risk tolerance scales the size of inventories, not the urgency with which traders trade, risk tolerance is intuitively proportional to assets under management. Our model thus captures an economically significant trade-off faced by large asset managers in financial markets.

Unlike Grossman and Stiglitz (1980) or Kyle (1985), our model has no noise traders or market makers. In the special case when traders believe other traders' signals are completely uninformative, the model implements the idea from Black (1986) of "trading on noise as if it were information." The speed of partial adjustment implements the intuition of Black (1995) concerning the "urgency" of trades. Valuation adjustments in traders' utility functions reflect an "appraisal premium" in the sense of Black and Treynor (1973).

In sharp contrast to the competitive model of Milgrom and Stokey (1982), our imperfectly competitive traders continue to trade on information after it has been incorporated into prices.

The information structure in which traders "agree to disagree" about the precision of their signals is the same as in Kyle and Lin (2001) and Scheinkman and Xiong (2003). The symmetry assumption prevents the number of state variables from exploding, avoiding the forecasting-the-forecasts-of-others problem described by Townsend (1983).

In our model, the price is a damped average of expectations based on common information but different prior distributions. Each trader believes that the time series of his own estimates of fundamental value has a martingale property implied by the law of iterated expectations. Each trader furthermore believes that the time series of other traders' estimates of fundamental value do not have a martingale property but instead exhibit mean reversion due to other traders' overconfidence. Since mean reversion dampens each trader's estimate of other traders' estimates, the equilibrium price, otherwise consistent with the beauty contest intuition of Keynes (1936), becomes a damped average of all traders' estimates. Kyle and Lin (2001) show that this dampening also occurs in a competitive model with the same information structure. Kyle, Obizhaeva and Wang (2013) explore implications for the equilibrium returns in more detail.

Allen, Morris and Shin (2006) describe the beauty contest of Keynes (1936) differently. They point out that the law of iterated expectations does not work for averages of expectations. Instead, the average expectation of the average expectation of fun-

damental value is dampened relative to the averaged expectation of fundamental value. Price fluctuations are dampened due to averaging expectations calculated with different information sets under a common prior. The models of Harris and Raviv (1993) and Banerjee, Kaniel and Kremer (2009) discuss how beauty contest intuition arises when traders process information differently.

We only consider equilibria in which the economy is always in a steady state. Describing how the equilibrium reaches a steady state, from a starting point in which prices are not already fully revealing, raises interesting issues related to Ostrovsky (2012), taking us beyond the scope of this paper.

Unlike Duffie (2010), flash crashes can result from fire sales not because capital has to move slowly but rather because traders—motivated by endogenously derived adverse-selection considerations—choose to move their capital slowly. The market microstructure invariants approach of Kyle and Obizhaeva (2013) also has implications for the relationship between trading costs and the speed with which capital moves.

Chan and Lakonishok (1995), Keim and Madhavan (1997), and Dufour and Engle (2000) provide empirical results consistent with the idea that the speed of trading influences price impact and execution costs. Holthausen, Leftwich and Mayers (1990) measure temporary and permanent price effects associated with block trades. They find that most of the adjustment occurs during the very first trade in a sequence, similar to the instantaneous price adjustment in our model. Almgren et al. (2005) calibrate price impact functions depending both on quantities traded and on the speed of trading, using a functional form very similar to the form we derive endogenously.

Numerous papers examine the economic implications of fast trading given exogenously specified price impact functions explicitly or implicitly depending on the speed of trading. Brunnermeier and Pedersen (2005) study price effects of a large trader unwinding his position in the presence of strategic traders; Carlin, Lobo and Viswanathan (2007) focus on the interactions among traders facing liquidity shocks; and Longstaff (2001) analyzes a portfolio choice problem; Grinold and Kahn (1995), Almgren and Chriss (2000), and Obizhaeva and Wang (2013) derive optimal execution strategies for liquidations of existing positions.

Our model not only makes the speed of trading endogenous but also suggests a specific functional form for temporary and permanent price impact consistent with existing theoretical and empirical literature on the subject.

This paper is structured as follows. Section I presents a one-period model. Section II outlines a dynamic continuous-time model. Section III examines properties of smooth trading equilibrium. Section IV concludes. Proofs are in the Appendix.

## I. One-period Model

To develop intuition for when an equilibrium exists and how equilibrium prices and quantities depend on the interaction between overconfidence and market power, we

start with a one-period model. There are  $N$  traders who trade a risky asset with liquidation value  $\tilde{v} \sim N(0, 1/\tau_v)$  against a safe numeraire asset with a liquidation value of one. Exponential utility traders with constant absolute risk aversion parameter  $A$  maximize their expected utility of terminal wealth. Each trader  $n$ ,  $n = 1, \dots, N$ , is endowed with an inventory of  $S_n$  shares of the risky asset. The risky asset is in zero net supply, implying that the sum of inventories of all traders is zero.

### Bayesian Updating.

All traders observe a public signal  $\tilde{i}_0 := \tau_0^{1/2} \cdot (\tau_v^{1/2} \tilde{v}) + \tilde{e}_0$  with  $\tilde{e}_0 \sim N(0, 1)$ . There are also  $N$  private signals  $\tilde{i}_n := \tau_n^{1/2} \cdot (\tau_v^{1/2} \tilde{v}) + \tilde{e}_n$  with  $\tilde{e}_n \sim N(0, 1)$ ,  $n = 1, \dots, N$ . The stock payoff  $\tilde{v}$ , the public signal error  $\tilde{e}_0$ , and  $N$  private signal errors  $\tilde{e}_1, \dots, \tilde{e}_N$  are independently distributed. Since  $\tau_v^{1/2} \tilde{v}$  and  $\tilde{e}$  have variances of one, the precision parameter  $\tau_n$  measures a signal-to-noise ratio. For the reasonable assumption that  $\tau_n$  is small, the  $R^2$  of a regression of  $\tau_v^{1/2} \tilde{v}$  on  $\tilde{i}_n$  is approximately  $\tau_n$  and the regression coefficient is approximately  $\tau_n^{1/2}$ . Trader  $n$  observes signal  $\tilde{i}_n$  privately but the equilibrium price, as discussed below, fully reveals the average of other traders' signals defined by  $\tilde{i}_{-n} := \frac{1}{N-1} \sum_{m \neq n} \tilde{i}_m$ .

Traders agree about the precision of the public signal  $\tau_0$  and agree to disagree about the precisions of private signals  $\tau_n$ . Traders are “relatively overconfident” in that each trader believes his own signal is more precise than the signals of the other traders; specifically, each trader  $n$  believes that  $\tau_n = \tau_H$  and  $\tau_m = \tau_L$  when  $m \neq n$ , with  $\tau_H > \tau_L \geq 0$ .<sup>2</sup>

Let  $E_n\{\dots\}$  and  $Var_n\{\dots\}$  denote trader  $n$ 's expectation and variance operators conditional on observing all signals  $i_0, i_1, \dots, i_N$ . Using standard formulas for conditional means and variances of jointly normally distributed random variables, we define

$$(1) \quad \tau := Var_n^{-1}\{\tilde{v}\} = \tau_v \cdot (1 + \tau_0 + \tau_H + (N - 1)\tau_L),$$

and obtain

$$(2) \quad E_n\{\tilde{v}\} = \frac{\tau_0^{1/2}}{\tau} \cdot \tilde{i}_0 + \frac{\tau_H^{1/2}}{\tau} \cdot \tilde{i}_n + \frac{(N - 1)\tau_L^{1/2}}{\tau} \cdot \tilde{i}_{-n}.$$

<sup>2</sup>For the case  $\tau_v = 1$  (without loss of generality), an earlier version of this paper defined the private signals as  $\tilde{i}_n = \tilde{v} + \tau_n^{-1/2} \cdot \tilde{e}_n$  instead of  $\tilde{i}_n = \tau_n^{1/2} \cdot \tilde{v} + \tilde{e}_n$ . More generally, the signal could be defined as  $\tilde{i}_n = K_n \cdot (\tau_n^{1/2} \cdot \tilde{v} + \tilde{e}_n)$ , with  $K_n = 1$  for the model in this paper and  $K_n = \tau_n^{-1/2}$  for the model in the previous version of this paper. It is a general principle of information processing that the scaling of the units in which information is measured does not affect conditional expectations. Nevertheless, for the scaling assumption  $K_n = \tau_n^{-1/2}$  made in the previous version of this paper, the scaling factor  $K_n = \tau_n^{-1/2}$  does change the equilibrium because traders agree to disagree about the value of  $K_n$ . We have chosen the scaling units  $K_n = 1$  in this version because the existence condition for the one-period model is the same as the existence condition derived for the continuous model. Furthermore, when  $K_n = 1$  and  $\tau_n$  is small, traders disagree about the correlation  $\tau_n^{1/2}/(1 + \tau_n)^{1/2}$  between  $\tilde{v}$  and  $\tilde{i}_n$  but do not disagree much about the variances  $Var\{\tilde{v}\} = 1$  and  $Var\{\tilde{i}_n\} = 1 + \tau_n$ . Developing this interesting issue further takes us beyond the scope of this paper.

### Utility Maximization with Market Power.

Traders are imperfect competitors who explicitly take into account the effect of their trading on prices. Suppose trader  $n$  believes the price is a function of the quantity  $x_n$  he trades,  $p = P(x_n)$ . He believes that his terminal wealth  $\tilde{W}_n := \tilde{v} \cdot (S_n + x_n) - P(x_n) \cdot x_n$  will be a normally distributed random variable with mean and variance given by

$$(3) \quad E_n\{\tilde{W}_n\} = E_n\{\tilde{v}\} \cdot (S_n + x_n) - P(x_n) \cdot x_n,$$

$$(4) \quad Var_n\{\tilde{W}_n\} = (S_n + x_n)^2 \cdot Var_n\{\tilde{v}\}.$$

Each trader  $n$  maximizes the exponential utility of his wealth,

$$(5) \quad E_n\{-e^{-A \cdot \tilde{W}_n}\} = -\exp\left(-A \cdot E_n\{\tilde{W}_n\} + \frac{1}{2} A^2 \cdot Var_n\{\tilde{W}_n\}\right).$$

The problem is equivalent to maximizing monotonically transformed expected utility  $-\frac{1}{A} \ln(-E_n\{-e^{-A \cdot \tilde{W}_n}\})$ . Plugging equations (1), (2), (3) and (4) into equation (5), the trader's optimization problem is to choose the quantity to trade  $x_n$  to solve the maximization problem

$$(6) \quad \max_{x_n} \left( \left[ \frac{\tau_0^{1/2}}{\tau} \cdot \tilde{i}_0 + \frac{\tau_H^{1/2}}{\tau} \cdot \tilde{i}_n + \frac{\tau_L^{1/2}}{\tau} \cdot (N-1) \cdot \tilde{i}_{-n} \right] \cdot (S_n + x_n) - P(x_n) \cdot x_n - \frac{A}{2\tau} \cdot (S_n + x_n)^2 \right).$$

For a perfect competitor,  $P(x_n)$  is a constant  $p$  which does not depend on  $x_n$ . In exercising market power, the oligopolistic trader takes into account how his choice of quantity  $x_n$  affects the price  $P(x_n)$ .

### Conjectured Linear Strategies.

Like Kyle (1989), we assume a single-price auction in which traders submit demand schedules  $X_n(i_0, i_n, S_n, p)$  to an auctioneer, who then calculates a market clearing price  $p$ . Suppose trader  $n$  conjectures that the other  $N-1$  traders submit symmetric linear demand schedules

$$(7) \quad X_m(i_0, i_m, S_m, p) = \alpha \cdot i_0 + \beta \cdot i_m - \gamma \cdot p - \delta \cdot S_m, \quad n = 1, \dots, N, \quad m \neq n.$$

From the market clearing condition  $\sum_{m=1}^N X_m(i_0, i_m, S_m, p) = 0$  and the linear specification of demand for the other traders, it follows that

$$(8) \quad x_n + \sum_{m \neq n} (\alpha \cdot i_0 + \beta \cdot i_m - \gamma \cdot p - \delta \cdot S_m) = 0.$$

Using zero-net-supply  $\sum_{m=1}^N S_m = 0$ , solving for  $p$  as a function of  $i_0$ ,  $i_{-n}$ ,  $S_n$ , and  $x_n$  yields the following linear residual demand schedule for trader  $n$ :

$$(9) \quad P(i_0, i_{-n}, S_n, x_n) = \frac{\alpha}{\gamma} \cdot i_0 + \frac{\beta}{\gamma} \cdot i_{-n} + \frac{1}{(N-1)\gamma} \cdot x_n + \frac{\delta}{(N-1)\gamma} \cdot S_n.$$

Under the tentative assumption that trader  $n$  knows the value of  $i_{-n}$ , we plug equation (9) into equation (6) and use the first order condition to find his optimal demand,

$$(10) \quad x_n = \frac{\left( \frac{\tau_0^{1/2}}{\tau} \cdot i_0 + \frac{\tau_H^{1/2}}{\tau} \cdot i_n + \frac{(N-1)\tau_L^{1/2}}{\tau} \cdot i_{-n} \right) - \left( \frac{\alpha}{\gamma} \cdot i_0 + \frac{\beta}{\gamma} \cdot i_{-n} \right) - \left( \frac{\delta}{(N-1)\gamma} + \frac{A}{\tau} \right) \cdot S_n}{\frac{2}{(N-1)\gamma} + \frac{A}{\tau}}.$$

In the numerator of this equation, the first term is trader  $n$ 's expectation of the liquidation value, the second term is the market clearing price when trader  $n$  trades a quantity of zero and has no inventory, and the last term is the adjustment for existing inventory. In the denominator, the first and second terms reflect how trader  $n$  restricts the quantity traded due to market power and risk aversion, respectively.

As in Kyle (1989), even though trader  $n$  does not observe  $i_{-n}$  explicitly, he is still able to implement this optimal strategy with a demand schedule which implicitly infers  $i_{-n}$  from the market clearing price; as in Kyle (1989), Rostek and Weretka (2012), and Du and Zhu (2013), the strategies are ex post optimal.

Define the constant  $C := 1/((N-1)\gamma) + A/\tau + \tau_L^{1/2}/(\tau\beta)$ . Solving for  $i_{-n}$  instead of  $p$  in the market clearing condition (8), substituting this solution into equation (9) above, and then solving for  $x_n$ , we derive a demand schedule  $X_n(i_0, i_n, S_n, p)$  for trader  $n$  as a function of price  $p$ ,

$$(11) \quad X_n(i_0, i_n, S_n, p) = \frac{1}{C} \cdot \left[ \left( \frac{\tau_0^{1/2}}{\tau} - \frac{(N-1)\tau_L^{1/2}}{\tau} \cdot \frac{\alpha}{\beta} \right) \cdot i_0 + \frac{\tau_H^{1/2}}{\tau} \cdot i_n + \left( \frac{(N-1)\tau_L^{1/2}}{\tau} \cdot \frac{\gamma}{\beta} - 1 \right) \cdot p - \left( \frac{\tau_L^{1/2}}{\tau} \cdot \frac{\delta}{\beta} + \frac{A}{\tau} \right) \cdot S_n \right].$$

### Equilibrium.

In a symmetric linear equilibrium, the strategy chosen by trader  $n$  must be the same as the linear strategy (7) conjectured for the other traders. Equating corresponding coefficients of variables  $i_0$ ,  $i_n$ ,  $P$  and  $S_n$  yields the system of four equations in terms of four unknowns  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$ :

$$(12) \quad \alpha = \frac{1}{C} \left( \frac{\tau_0^{1/2}}{\tau} - \frac{(N-1)\tau_L^{1/2}}{\tau} \cdot \frac{\alpha}{\beta} \right), \quad \beta = \frac{1}{C} \frac{\tau_H^{1/2}}{\tau},$$

$$(13) \quad \gamma = -\frac{1}{C} \left( \frac{(N-1)\tau_L^{1/2}}{\tau} \cdot \frac{\gamma}{\beta} - 1 \right), \quad \delta = \frac{1}{C} \left( \frac{\tau_L^{1/2}}{\tau} \cdot \frac{\delta}{\beta} + \frac{A}{\tau} \right).$$

The unique solution is

$$(14) \quad \beta = \frac{(N-2)\tau_H^{1/2} - 2(N-1)\tau_L^{1/2}}{A \cdot (N-1)},$$

$$\alpha = \frac{\tau_0^{1/2}}{\tau_H^{1/2} + (N-1)\tau_L^{1/2}} \cdot \beta, \quad \gamma = \frac{\tau}{\tau_H^{1/2} + (N-1)\tau_L^{1/2}} \cdot \beta, \quad \delta = \frac{A}{\tau_H^{1/2} - \tau_L^{1/2}} \cdot \beta.$$

**THEOREM 1:** Define the constant  $\Delta_H := \tau_H^{1/2} - 2 \cdot (\tau_H^{1/2} + (N-1)\tau_L^{1/2})/N$ . In addition to a no-trade equilibrium, which always exists, there exists a unique symmetric equilibrium with linear trading strategies and non-zero trade if and only if  $\Delta_H > 0$ . Such an equilibrium has the following properties:

1. The equilibrium demand functions are given by equations (7) and (14);
2. The equilibrium quantity traded by trader  $n$  is

$$(15) \quad x_n^* = \frac{\Delta_H}{A} \cdot (\tilde{i}_n - \tilde{i}_{-n}) - \delta \cdot S_n;$$

3. The equilibrium price is

$$(16) \quad P^* = \frac{\tau_0^{1/2}}{\tau} \cdot \tilde{i}_0 + \frac{\tau_H^{1/2} + (N-1)\tau_L^{1/2}}{\tau} \cdot \frac{1}{N} \sum_{m=1}^N \tilde{i}_m.$$

The second order condition for optimization (6) is equivalent to the denominator of equation (10) being positive, i.e.,  $\frac{2}{(N-1)\gamma} + \frac{A}{\tau} > 0$ . Given the solution for  $\gamma$  in equation (14), this second order condition holds if and only if  $\Delta_H > 0$ . From equation (14), the condition  $\Delta_H > 0$  also ensures that a trader's demand function is increasing in the trader's own private signal ( $\beta > 0$ ), increasing in the public signal ( $\alpha > 0$ ), decreasing in price ( $\gamma > 0$ ), and decreasing in the trader's inventory ( $\delta > 0$ ). From equation (15), the trader trades in the direction of his private signal and against the average of the signals of others.

We think of  $\Delta_H$  as a measure of “disagreement.” The condition  $\Delta_H > 0$  is equivalent to

$$(17) \quad \frac{\tau_H^{1/2}}{\tau_L^{1/2}} > 2 + \frac{2}{N-2}.$$

To obtain an equilibrium with positive trading volume, there needs to be “enough” disagreement. A symmetric linear equilibrium with trade does not exist unless  $N \geq 3$  and  $\tau_H^{1/2}$  is sufficiently more than twice as large as  $\tau_L^{1/2}$ .

As in the models of Kyle (1989) and Rostek and Weretka (2012), traders “shade” their bids by trading approximately one-half of the amount they believe would fully reveal their private information. Intuitively, if there is not enough disagreement to sustain an equilibrium with trade, each trader wants to shade his bid more than the others, and the result is no trade.<sup>3</sup>

<sup>3</sup>When there is not enough disagreement to sustain an equilibrium with pure strategies, one might imagine that it is possible to have an equilibrium with mixed strategies. For mixed strategies

## Equilibrium Properties.

The fully revealing equilibrium price (16) is the average of all traders' risk-neutral buy-and-hold valuations of the risky asset. The price is a weighted average of square roots of precisions, with the weight on the public signal  $i_0$  proportional to  $\tau_0^{1/2}$  and the weight on the  $N$  private signals  $i_n$  proportional to the weighted average of the square roots of the high and low precisions  $\tau_H$  and  $\tau_L$ , i.e.,  $[\tau_H^{1/2} + (N - 1)\tau_L^{1/2}]/N$ . Trade occurs because each trader believes that the market price weights the signals incorrectly. Each trader believes that his own signal should receive a higher weight  $\tau_H^{1/2}$  and the other  $N - 1$  signals should each receive a lower weight  $\tau_L^{1/2}$ . There is no risk adjustment, because the risky asset is in zero-net supply and there are no noise traders.

Traders trade for both information and hedging motives. The equilibrium quantity traded  $x_n^*$  in equation (15) is a linear function of a trader's inventory  $S_n$  and the deviation of the trader's signal from the average of other traders' signals  $(\tilde{i}_n - \frac{1}{N} \sum_{m \neq n} \tilde{i}_m)$ .

Each trader "shades" the quantity traded relative to the quantity that a perfect competitor would trade to exercise his market power. To quantify this bid shading, define a trader's "target inventory"  $S_n^{TI}$  as the inventory such that he does not want to trade ( $x_n^* = 0$ ), given from equation (15) by

$$(18) \quad S_n^{TI} = \frac{1}{A} \cdot \left(1 - \frac{1}{N}\right) \cdot (\tau_H^{1/2} - \tau_L^{1/2}) \cdot (\tilde{i}_n - \tilde{i}_{-n}).$$

Then, his optimal demand can be written

$$(19) \quad x_n^* = \delta \cdot (S_n^{TI} - S_n),$$

where the parameter  $\delta$  is defined in equation (14).

These equations have a simple intuition. A target inventory  $S_n^{TI}$  is proportional to (1) a trader's risk tolerance  $1/A$ ; (2) the difference between his valuation and the valuation of other traders, which itself is proportional to overconfidence  $\tau_H^{1/2} - \tau_L^{1/2}$ ; and (3) the difference between a trader's signal and the average signal of others  $(\tilde{i}_n - \tilde{i}_{-n})$ . Even if  $N$  goes to infinity, traders continue to agree to disagree, and their target inventories do not converge to zero.

to be an equilibrium, the trader must be indifferent across the various randomized choices of quantities he trades. For example, if we add normally distributed noise to quantities traded, symmetrically across all traders, a mixed strategy equilibrium requires the second order condition to be exactly zero. This means that the quadratic objective function reduces to a linear function, i.e., the denominator of equation (10) is zero. Since the trader has to be indifferent across various randomizations, this further implies that the linear function must be a constant, independently of the quantity traded. This assumption cannot hold, because a trader with a positive value of  $i_n$  would always want to buy unlimited quantities, and a trader with a negative  $i_n$  would always want to sell unlimited quantities. Thus, an equilibrium with symmetric normally distributed noise cannot exist. When noise is not normally distributed or the equilibrium is not symmetric, the objective function is not quadratic any more, but it will still be difficult to find a mixed strategy equilibrium given that the sensitivity of utility to a trader's own private information must be well-defined.

The parameter  $\delta$  determines the fraction by which traders adjust positions towards target levels. This parameter is always less than one, decreasing monotonically in precision  $\tau_L^{1/2}$  from  $\delta = 1 - 1/(N-1) < 1$  when  $\tau_L^{1/2} = 0$  to  $\delta = 0$  when  $\tau_L^{1/2} = (1 - 1/(N-1)) \cdot \tau_H^{1/2}/2$ , which corresponds to  $\Delta_H = 0$ . If traders were perfect competitors, it could be shown that the competitive equilibrium price would still be defined by equation (16), but the optimal demand would equal  $x_n^* = S_n^{TI} - S_n$ , i.e., traders would move all the way from initial inventory  $S_n$  to target inventory  $S_n^{TI}$  with  $\delta = 1$  in equation (19). Exercise of monopoly power reduces the amount of trading relative to perfect competition.

Equation (9) implies that the price impact function has the form

$$(20) \quad P(x_n, S_n) = \lambda_0 + \lambda_S \cdot S_n + \lambda_x \cdot x_n.$$

Using equations (13) and (14), its coefficients are

$$(21) \quad \lambda_S := \frac{\delta}{(N-1)\gamma} = \frac{\tau_H^{1/2} + (N-1)\tau_L^{1/2}}{\tau} \cdot \frac{A}{(N-1)(\tau_H^{1/2} - \tau_L^{1/2})}.$$

and

$$(22) \quad \lambda_x := \frac{\lambda_S}{\delta} = \frac{1}{(N-1)\gamma} = \frac{A}{N \cdot \Delta_H} \cdot \frac{\tau_H^{1/2} + (N-1)\tau_L^{1/2}}{\tau}.$$

In equilibrium, traders provide liquidity to one another because they agree to disagree about the quality of their respective signals. The liquidity measures  $\lambda_S$  and  $\lambda_x$  depend on the degree of disagreement and on risk aversion. In the continuous-time model, which we consider next, the first component  $\lambda_S \cdot S_n$  will be related to permanent linear impact as in Kyle (1985). The second component  $\lambda_x \cdot x_n$  will be related to temporary price impact determined by the speed of trading, with  $x_n^*$  replaced by the derivative of the trader's inventory  $dS_n/dt$ . As we shall see below, the continuous-time model sharpens the insights derived from this one-period model.

## II. Continuous-time Model

The continuous-time model has the following structure.

There are  $N$  risk-averse oligopolistic traders who trade at price  $P(t)$  a risky asset in zero net supply against a risk-free asset which earns constant risk-free rate  $r > 0$ .

The risky asset pays out dividends at continuous rate  $D(t)$ . Dividends follow a stochastic process with mean-reverting stochastic growth rate  $G^*(t)$ , constant instantaneous volatility  $\sigma_D > 0$ , and constant rate of mean reversion  $\alpha_D > 0$ ,

$$(23) \quad dD(t) := -\alpha_D \cdot D(t) \cdot dt + G^*(t) \cdot dt + \sigma_D \cdot dB_D(t).$$

The growth rate  $G^*(t)$  follows an AR-1 process with mean reversion  $\alpha_G$  and volatility  $\sigma_G$ :

$$(24) \quad dG^*(t) := -\alpha_G \cdot G^*(t) \cdot dt + \sigma_G \cdot dB_G(t).$$

The dividend is publicly observable, but the growth rate  $G^*(t)$  is not observed by any trader.

Each trader  $n$  observes a continuous stream of private information  $I_n(t)$  defined by the stochastic process

$$(25) \quad dI_n(t) := \tau_n^{1/2} \cdot \frac{G^*(t)}{\sigma_G \cdot \Omega^{1/2}} \cdot dt + dB_n(t), \quad n = 1, \dots, N.$$

Since its drift  $\tau_n^{1/2} \cdot G^*(t)/(\sigma_G \Omega^{1/2})$  is proportional to  $G^*(t)$ , each increment  $dI_n(t)$  in the process  $I_n(t)$  is a noisy observation of the unobserved growth rate  $G^*(t)$ . In equation (25), the parameters  $\sigma_G$  and  $\Omega$  are scaling parameters which simplify the intuitive interpretations of the model. The “precision” parameter  $\tau_n$  measures the informativeness of the signal  $dI_n(t)$  as a signal-to-noise ratio describing how fast the information flow generates a signal of a given level of statistical significance. The parameter  $\Omega$  measures the steady-state error variance in units of time, as discussed below.<sup>4</sup>

Analogous to the one-period model, each trader believes that his own private information  $I_n(t)$  has “high” precision  $\tau_n = \tau_H$  and the other traders’ private information has “low” precision  $\tau_m = \tau_L$  for  $m \neq n$ , with  $\tau_H > \tau_L \geq 0$ .

Each trader’s information set at time  $t$  consists of the histories of the dividend process  $D(s)$ , the trader’s own private information  $I_n(s)$ , and the market price  $P(s)$ ,  $s \in [-\infty, t]$ . All traders process information rationally; they apply Bayes Law correctly given their possibly incorrect beliefs.

Using the scaling parameter  $\Omega$ , we can express the information content of the publicly observable dividend  $D(t)$  in a form consistent with the notation for private signals  $I_n(t)$  in equation (25). Define  $dI_0(t) := [\alpha_D \cdot D(t) \cdot dt + dD(t)] / \sigma_D$  and  $\tau_0 := \Omega \cdot \sigma_G^2 / \sigma_D^2$  with  $dB_0 := dB_D$ . Then the public information  $I_0(t)$  in the divided stream (23) can be written

$$(26) \quad dI_0(t) := \tau_0^{1/2} \cdot \frac{G^*(t)}{\sigma_G \cdot \Omega^{1/2}} \cdot dt + dB_0(t).$$

Observing the process  $I_0(t)$  is informationally equivalent to observing the dividend process. The quantity  $\tau_0$  measures the precision of the dividend process in units analogous to the units of precision for private signals. This notation simplifies the Kalman filtering formulas we are about to derive.

The values of the parameters  $\alpha_D$ ,  $\sigma_D$ ,  $\alpha_G$ ,  $\sigma_G$ ,  $\tau_H$ ,  $\tau_L$ , and  $\Omega$  are common knowledge. It is common knowledge that  $B_D(t)$ ,  $B_G(t)$ ,  $B_1(t), \dots, B_N(t)$  are independently distributed standardized Brownian motions. “Relatively overconfident” traders stubbornly believe that their beliefs  $\tau_H$  or  $\tau_L$  about the precision parameters  $\tau_0, \dots, \tau_N$

<sup>4</sup>Since the innovation variance of the signal  $dI_n(t)$  can be estimated arbitrarily precisely by observing past signals continuously, traders “agree to agree” that the innovation variance of the signal is one. Scaling the innovation variance of  $I_n(t)$  in equation (25) to make it equal one is therefore a normalization without loss of generality. The continuous-time model is different from the one-period model (see footnote 2).

are point estimates with no possibility of error. This stubborn belief structure is also common knowledge. Traders thus “agree to disagree” about the precisions of their private signals.<sup>5</sup>

Let  $S_n(t)$  denote the inventory of trader  $n$  at time  $t$ . Since the risky asset is in zero net supply, we have  $\sum_{n=1}^N S_n(t, P(t)) = 0$ . We conjecture that all traders smooth out their trading over time, i.e., the trajectories of their inventories  $S_n(t)$  are differentiable functions of time. Intuitively, infinitely fast portfolio updating cannot be an equilibrium. If other traders traded infinitely fast, each trader would then believe that he could lower his execution costs by trading more slowly than the other traders—essentially by walking up or down the residual demand schedules they present to him—but all traders cannot trade more slowly than average. For example, in Kyle (1985) it turns out to be optimal for the informed trader to smooth out his trading so that his inventory is a continuous function of time. This is consistent with equilibrium because the noise traders in Kyle (1985) do not trade optimally; they generate high transactions costs by trading infinitely impatiently, so that their inventories are diffusion, not a differentiable function of time. If the noise traders in Kyle (1985) were modeled as rational hedgers motivated by endowment shocks, they would smooth their trading, as in Vayanos (1999), and this would “break” the equilibrium in Kyle (1985).

We therefore specify trading strategies and market clearing condition in terms of rates of trading, not shares traded. Each trader’s trading strategy is assumed to be a mapping from his information set at time  $t$  into a “flow demand schedule”  $X_n(t, .)$  defining his “trading intensity” as a function of its price  $P(t)$ . An auctioneer continuously calculates the market clearing price  $P(t)$  such that  $\sum_{n=1}^N X_n(t, P(t)) = 0$ . The trader’s inventory follows  $dS_n(t)/dt = X_n(t, P(t))$ . Each trader takes into account the effect of his trading on market prices.

Each trader chooses a consumption intensity  $c_n(t)$  and trading strategy  $X_n(t, .)$  to maximize an expected constant-absolute-risk-aversion (CARA) utility function. Let  $U(c_n(s)) := -e^{-A c_n(s)}$  be an exponential utility function with a constant absolute risk aversion parameter  $A$ . Letting  $\rho > 0$  denote a time preference parameter, trader  $n$  solves the maximization problem

$$(27) \quad \max_{\{c_n(t), X_n(t, .)\}} E_t^n \left\{ \int_{s=t}^{\infty} e^{-\rho(s-t)} \cdot U(c_n(s)) \cdot ds \right\},$$

subject to the inventory constraint  $dS_n(t) = X_n(t, P(t)) \cdot dt$  and the budget constraint  $dW = (r \cdot W(t) + S_n(t) \cdot D(t) - c_n(t) - P(t) \cdot X_n(t, P(t))) \cdot dt$ . In the notation  $E_t^n \{ \dots \}$ , the superscript  $n$  indicates that the expectation is taken with respect to the beliefs of trader  $n$ ; The subscript  $t$  indicates that the expectation is taken with

<sup>5</sup>We call this belief structure “relative overconfidence” to distinguish it from a belief structure with “absolute overconfidence” in which traders believe the precisions of their signals are greater than empirically true precisions. Empirically true precisions do not affect the equilibrium strategies investigated in this paper but do affect empirical predictions about asset returns, which takes us beyond the scope of this paper.

respect to trader  $n$ 's information set at time  $t$ . As shown below, in equilibrium, each trader can infer the average of other traders' private signals from the history of prices, so all traders act as if they are fully informed with the same information  $D(t), I_1(t), \dots, I_N(t)$ .

We will show numerically that if disagreement is large enough—i.e., if  $\tau_H$  is sufficiently larger than  $\tau_L$ —there will be trade based on private information. The perceived precisions  $\tau_L$  and  $\tau_H$  affect the equilibrium prices and quantities traded. Without overconfidence—e.g., in a model with rational expectations—there would be no trade after traders unwind their suboptimal initial endowments.

### Bayesian Updating with Signals of Arbitrary Precision.

In general, discussing how beliefs of traders about the information content of signals affect the information content of prices is tricky because the discussion requires notation which keeps track of the unobserved true values of the parameters, the beliefs of an economist who studies the market outcomes, and the possibly incorrect beliefs of the traders in the market. A hypothetical economist who studies this equilibrium may assume precisions arbitrarily different from the traders in the market, and the traders in the market may, in principle, assume values arbitrarily different from other traders. Of course, by studying the equilibrium of the economy, an economist will not change the equilibrium, but the economist's beliefs about the values of the precisions will affect how the economist interprets the information content of prices.

We therefore first study information processing for arbitrary “generic” beliefs  $\bar{\tau}_0, \bar{\tau}_1, \dots, \bar{\tau}_N$  about the precisions.

Define  $G(t) = E_t\{G^*(t)\}$ , where the subscript  $t$  denotes conditioning on the history of the signals  $I_0(s), \dots, I_N(s)$  for  $s \in [-\infty, t]$ . Without loss of generality, we define  $\bar{\Omega}$  as the error variance  $\bar{\Omega} := \text{Var}\{(G^*(t) - G(t))/\sigma_G\}$ . We assume a steady state in which  $\bar{\Omega}$  is a constant which does not depend on time. Like a squared Sharpe ratio,  $\bar{\Omega}$  measures the error variance in units of time. For example, if time is measured in years,  $\bar{\Omega} = 4$  means that the estimate of  $G^*(t)$  is “behind” the true value of  $G^*(t)$  by an amount equivalent to four years of volatility unfolding at rate  $\sigma_G$ . There are simple and intuitive formulas for information processing:

LEMMA 1: *Given generic beliefs  $\bar{\tau}_1, \dots, \bar{\tau}_N$ , let  $\bar{\tau}$  denote the sum of precisions*

$$(28) \quad \bar{\tau} := \bar{\tau}_0 + \sum_{n=1}^N \bar{\tau}_n.$$

*Then  $\bar{\Omega}$  and  $dG(t)$  satisfy*

$$(29) \quad \bar{\Omega}^{-1} := \text{Var}^{-1} \left\{ \frac{G^*(t) - G(t)}{\sigma_G} \right\} = 2 \cdot \alpha_G + \bar{\tau},$$

$$(30) \quad dG(t) = -(\alpha_G + \bar{\tau}) \cdot G(t) \cdot dt + \sigma_G \cdot \bar{\Omega}^{1/2} \cdot \sum_{n=0}^N \bar{\tau}_n^{1/2} \cdot dI_n.$$

The error variance  $\bar{\Omega}$  corresponds to a steady state that balances an increase in error variance due to stochastic change  $dB_G(t)$  in the true growth rate with a reduction in error variance due to a mean-reversion of the true growth rate at rate  $\alpha_G$  and arrival of new information with total precision  $\bar{\tau}$ .

Note that  $\bar{\Omega}$  is not a “free parameter,” but is instead determined as an endogenous function of the other parameters. Equation (29) implies that  $\Omega$  turns out to be the solution to the quadratic equation  $\bar{\Omega}^{-1} = 2 \cdot \alpha_G + \bar{\Omega} \cdot \sigma_G^2 / \sigma_D^2 + \sum_{n=1}^N \bar{\tau}_n$ . In equations (25) and (26), we scaled the units with which precision is measured by the endogenous parameter  $\Omega$  because this leads to simpler Kalman filtering expressions which more clearly bring out intuition about signal processing.

From equation (30), the estimate  $G(t)$  can be conveniently written as the weighted sum of  $N + 1$  sufficient statistics  $H_n$  corresponding to information flow  $dI_n$ . Define the sufficient statistics  $H_n(t)$  by

$$(31) \quad H_n(t) := \int_{u=-\infty}^t e^{-(\alpha_G + \bar{\tau}) \cdot (t-u)} \cdot dI_n(u), \quad n = 0, 1, \dots, N,$$

which implies

$$(32) \quad dH_n(t) = -(\alpha_G + \bar{\tau}) \cdot H_n(t) \cdot dt + dI_n(t), \quad n = 0, 1, \dots, N.$$

Then  $G(t)$  becomes a linear combination of sufficient statistics  $H_n(t)$  with weights proportional to the square roots of the precisions  $\bar{\tau}_n^{1/2}$ :

$$(33) \quad G(t) = \sigma_G \cdot \bar{\Omega}^{1/2} \cdot \sum_{n=0}^N \bar{\tau}_n^{1/2} \cdot H_n(t).$$

The importance of each bit of information  $dI_n$  about the growth rate  $G(t)$  decays exponentially at a rate  $\alpha_G + \bar{\tau}$ , which is the same for all of the signals. The half-life of a signal  $\ln 2 / (\alpha_G + \bar{\tau})$  decreases as the “aggregate precision”  $\bar{\tau}$  increases. Even though the true unobserved growth rate may have a long half life (small  $\alpha_G$ ), information about this growth rate may decay rapidly if aggregate precision  $\bar{\tau}$  is large.

Note that equations (25), (26), and (30) imply that the estimate  $G(t)$  mean-reverts to zero at a rate  $\alpha_G$  while moving towards the true value  $G^*$  at rate  $\bar{\tau}$ :

$$(34) \quad dG(t) = -\alpha_G \cdot G(t) \cdot dt + \bar{\tau} \cdot (G^* - G) \cdot dt + \sigma_G \cdot \bar{\Omega}^{1/2} \cdot \sum_{n=0}^N \bar{\tau}_n^{1/2} \cdot dB_n(t).$$

### Bayesian Updating by Traders in the Model.

Asset managers who trade based on statistical models typically take raw information and process it into “signals.” Returns forecasts are then generated as functions

of the signals. The signals are often scaled so that they have a meaningful interpretation in terms of intuition or statistics. Here, we can think of the information processes  $I_n(t)$  as “raw information” and the sufficient statistics  $H_n(t)$  as “signals.” In equilibrium, traders believe that their signals forecast returns. Given the depth of the market, they trade on the signals with an aggressiveness that depends on the information content and the decay rate of the signals.

We next consider how traders “in the model” use their signals to update their beliefs about the unobserved growth rate  $G^*(t)$ .

Let  $G_n(t) := E_t^n\{G^*(t)\}$  denote trader  $n$ ’s estimate of the unobserved growth rate  $G^*(t)$  conditional on all information. The superscript  $n$  indicates that conditional distributions of growth rates are calculated by trader  $n$  based on his belief that his own signal has high precision  $\tau_H$  and other traders’ signals have low precision  $\tau_L$ . The subscript  $t$  denotes, as before, conditioning on the history of all information  $I_0(s) \equiv D(t), I_1(s), \dots, I_N(s), s \in [-\infty, t]$ .

It is common knowledge that each trader believes his own signal has high precision  $\tau_H$  and other traders’ signals have low precision  $\tau_L$ . Thus, if we define

$$(35) \quad \tau := \tau_0 + \tau_H + (N - 1)\tau_L, \quad \Omega^{-1} := 2\alpha_G + \tau,$$

all traders agree that the error variance is given by  $\bar{\Omega} = \Omega$  from equation (29), total precision is given by  $\bar{\tau} = \tau$  from equation (28), with  $\tau_0 = \Omega \cdot \sigma_G^2 / \sigma_D^2$ . Traders agree that the correct way to process available information is to construct signals  $H_n(t), n = 0, \dots, N$  by plugging  $\tau$  and  $\Omega$  into equation (31) and (32). Because they disagree about the precisions, traders disagree about the weights used to aggregate the signals  $H_n(t), n = 0, \dots, N$  into an estimate of a growth rate in equation (33); each assigns a larger weight to his own signal than to others’ signals.

Let  $H_{-n}(t)$  denote the average of the other traders’ signals  $m \neq n$ :

$$(36) \quad H_{-n}(t) := \frac{1}{N-1} \sum_{m=1, \dots, N; m \neq n} H_m(t).$$

Equation (33) implies that trader  $n$ ’s estimate of the true growth rate  $G_n(t)$  can be expressed as a linear combination of three signals  $H_0(t)$ ,  $H_n(t)$ , and  $H_{-n}(t)$ :

$$(37) \quad G_n(t) := \sigma_G \cdot \Omega^{1/2} \cdot \left( \tau_0^{1/2} \cdot H_0(t) + \tau_H^{1/2} \cdot H_n(t) + (N - 1)\tau_L^{1/2} \cdot H_{-n}(t) \right).$$

Trader  $n$ ’s optimal trading strategy depends on trader  $n$ ’s estimates of the unobserved growth rate  $G^*(t)$  and his beliefs about the dynamic statistical relationship between this rate and the signals  $H_0(t)$ ,  $H_n(t)$  and  $H_{-n}(t)$ .

Define the  $N + 1$  processes  $dB_0^n$ ,  $dB_n^n$ , and  $dB_m^n$ ,  $m = 1, \dots, N$ ,  $m \neq n$ , by

$$(38) \quad dB_0^n(t) = \tau_0^{1/2}(\sigma_G \Omega^{1/2})^{-1} \cdot (G^*(t) - G_n(t)) \cdot dt + dB_D(t),$$

$$(39) \quad dB_n^n(t) = \tau_H^{1/2}(\sigma_G \Omega^{1/2})^{-1} \cdot (G^*(t) - G_n(t)) \cdot dt + dB_n(t),$$

and

$$(40) \quad dB_m^n(t) = \tau_L^{1/2} (\sigma_G \Omega^{1/2})^{-1} \cdot (G^*(t) - G_n(t)) \cdot dt + dB_m(t).$$

The superscript  $n$  indicates conditioning on beliefs of trader  $n$ . Since trader  $n$ 's forecast of the error  $G^*(t) - G_n(t)$  is zero given his information set, these  $N + 1$  processes are independently distributed Brownian motions from the perspective of trader  $n$ . In terms of these Brownian motions, trader  $n$  thinks that signals change as follows:

$$(41) \quad dH_0(t) = -(\alpha_G + \tau) \cdot H_0(t) \cdot dt + \tau_0^{1/2} \cdot (\sigma_G \Omega^{1/2})^{-1} \cdot G_n(t) \cdot dt + dB_0^n(t),$$

$$(42) \quad dH_n(t) = -(\alpha_G + \tau) \cdot H_n(t) \cdot dt + \tau_H^{1/2} \cdot (\sigma_G \Omega^{1/2})^{-1} \cdot G_n(t) \cdot dt + dB_n^n(t),$$

$$(43) \quad dH_{-n}(t) = -(\alpha_G + \tau) \cdot H_{-n}(t) \cdot dt + \tau_L^{1/2} \cdot (\sigma_G \Omega^{1/2})^{-1} \cdot G_n(t) \cdot dt + \frac{1}{N-1} \sum_{m=1,..N; m \neq n} dB_m^n(t).$$

Note that each signal drifts towards zero at rate  $\alpha_G + \tau$  and drifts towards the optimal forecast  $G_n(t)$  at a rate proportional to the square root of the signal's precisions  $\tau_0^{1/2}$ ,  $\tau_H^{1/2}$ , or  $\tau_L^{1/2}$ , respectively.

### Utility Maximization with Market Power.

We conjecture a steady state value function  $V(M_n, S_n, D, H_0, H_n, H_{-n})$ , where  $M_n$  denotes trader  $n$ 's cash holdings (measured in dollars) and  $S_n$  denotes trader  $n$ 's holdings of the traded asset (measured in shares).

In a competitive model, a trader's value function depends on his wealth but does not depend on the decomposition of his wealth into his various security holdings. With imperfect competition, the decomposition of a trader's wealth into various security holdings does affect his value function, because the trader cannot costlessly convert one security holding into cash or another security holding by trading at market prices. "Wealth" does not appear in the value function because wealth is not well-defined. Trader  $n$  is always influencing the mark-to-market value of his risky inventory by choosing his rate of trading. It is therefore necessary to keep track of the two components of wealth—cash  $M_n$  and inventories  $S_n$ —separately.

Also, we expect the asset price to be a linear combination of two components: (1) a dividend level component linear in dividends  $D(t)$  (measured in dollars per share per unit of time) and (2) a dividend-growth component linear in the variables  $H_0(t)$ ,  $H_n(t)$ , and  $H_{-n}(t)$ . As in the one-period model, we use the concept of "no-regret" pricing, based on the intuition that by conditioning on the market price, trader  $n$  can achieve the same outcome that could be obtained if he directly observed the average of other traders' signals  $H_{-n}(t)$ . Therefore we include  $H_{-n}(t)$  as a state variable in the value function and omit the price  $P(t)$ .

In deriving the equilibrium below, the problem is simplified if the three state variables  $H_0(t)$ ,  $H_n(t)$ , and  $H_{-n}(t)$  are replaced with two “composite” signals, which we denote  $\hat{H}_n(t)$  and  $\hat{H}_{-n}(t)$ . Define the weighting constant  $\hat{A}$  by

$$(44) \quad \hat{A} := \frac{\tau_0^{1/2}}{\tau_H^{1/2} + (N-1)\tau_L^{1/2}}.$$

Now define the two composite signals  $\hat{H}_n(t)$  and  $\hat{H}_{-n}(t)$  by

$$(45) \quad \hat{H}_n(t) := H_n(t) + \hat{A} \cdot H_0(t),$$

$$(46) \quad \hat{H}_{-n}(t) := H_{-n}(t) + \hat{A} \cdot H_0(t).$$

Trader  $n$ 's estimate of dividend growth rate can now be expressed as a function of the two composite signals  $\hat{H}_n(t)$  and  $\hat{H}_{-n}(t)$  as

$$(47) \quad G_n(t) = \sigma_G \cdot \Omega^{1/2} \left( \tau_H^{1/2} \cdot \hat{H}_n(t) + (N-1)\tau_L^{1/2} \cdot \hat{H}_{-n}(t) \right).$$

In terms of the composite variables  $\hat{H}_n$  and  $\hat{H}_{-n}$ , we conjecture (and verify below) a steady state value function of the form  $V(M_n, S_n, D, \hat{H}_n, \hat{H}_{-n})$ . Letting  $(c_n(t), X_n(t, \cdot))$  denote the optimal consumption and investment policy, we have

$$(48) \quad V(M, S_n, D, \hat{H}_n, \hat{H}_{-n}) := \max_{\{c_n(t), X_n(t, \cdot)\}} E_t^n \left\{ \int_{s=t}^{\infty} -e^{-\rho(s-t)-A \cdot c_n(s)} \cdot ds \right\}.$$

The five state variables satisfy five stochastic differential equations

$$(49) \quad dM_n(t) = (r \cdot M_n(t) + S_n(t) \cdot D(t) - c_n(t) - P(x_t) \cdot X_n(t, P(t))) \cdot dt,$$

$$(50) \quad dS_n(t) = X_n(t, P(t)) \cdot dt,$$

$$(51) \quad dD(t) = -\alpha_D \cdot D(t) \cdot dt + G_n(t) \cdot dt + \sigma_D \cdot dB_0^n(t),$$

$$(52) \quad \begin{aligned} d\hat{H}_n(t) = & -(\alpha_G + \tau) \cdot \hat{H}_n(t) \cdot dt \\ & + (\tau_H^{1/2} + \hat{A}\tau_0^{1/2}) \cdot \left( \tau_H^{1/2} \cdot \hat{H}_n(t) + (N-1)\tau_L^{1/2} \cdot \hat{H}_{-n}(t) \right) \cdot dt \\ & + \hat{A} \cdot dB_0^n(t) + dB_n^n(t), \end{aligned}$$

$$(53) \quad \begin{aligned} d\hat{H}_{-n}(t) = & -(\alpha_G + \tau) \cdot \hat{H}_{-n}(t) \cdot dt \\ & + (\tau_L^{1/2} + \hat{A}\tau_0^{1/2}) \cdot \left( \tau_H^{1/2} \cdot \hat{H}_n(t) + (N-1)\tau_L^{1/2} \cdot \hat{H}_{-n}(t) \right) \cdot dt \\ & + \hat{A} \cdot dB_0^n(t) + \frac{1}{N-1} \sum_{m=1,..N; m \neq n} dB_m^n(t). \end{aligned}$$

The dynamics of  $\hat{H}_n$  and  $\hat{H}_{-n}$  in equations (52) and (53) can be derived from equations (41), (42), and (43).

The value function  $V(\cdot)$  satisfies the transversality condition

$$(54) \quad \lim_{t \rightarrow +\infty} E^n \{e^{-\rho t} V(M_n(t), S_n(t), D(t), \hat{H}_n(t), \hat{H}_{-n}(t))\} = 0.$$

### Linear Conjectured Strategies.

Based on his information set, each trader submits a flow demand schedule for the rate at which he will buy the asset at time  $t$  as a function of the market clearing price. Trader  $n$  conjectures that the other  $N - 1$  traders,  $m = 1, \dots, N, m \neq n$ , submit symmetric linear demand schedules of the form

$$(55) \quad X_m(t) = dS_n(t)/dt = \gamma_D \cdot D(t) + \gamma_H \cdot \hat{H}_m(t) - \gamma_S \cdot S_m(t) - \gamma_P \cdot P(t).$$

The demand schedules are defined by the four constants  $\gamma_D$ ,  $\gamma_H$ ,  $\gamma_S$ , and  $\gamma_P$ .

Let  $x_n(t) = X_n(t, P(t)) = dS_n(t)/dt$  denote the “flow-quantity” traded by trader  $n$ . From the market clearing condition and the linear conjecture for demand schedules of other traders, it follows that

$$(56) \quad x_n(t) + \sum_{m=1,..N; m \neq n} \left( \gamma_D \cdot D(t) + \gamma_H \cdot \hat{H}_m(t) - \gamma_S \cdot S_m(t) - \gamma_P \cdot P(t) \right) = 0.$$

Since zero net supply implies  $\sum_{m=1}^N S_m(t) = 0$ , solving for  $P(t)$  as a function of  $x_n(t)$  yields the following price impact function that trader  $n$  conjectures he faces:

$$(57) \quad P(x_n(t)) = \frac{\gamma_D}{\gamma_P} \cdot D(t) + \frac{\gamma_H}{\gamma_P} \cdot \hat{H}_{-n}(t) + \frac{\gamma_S}{\gamma_P} \frac{1}{N-1} \cdot S_n(t) + \frac{1}{(N-1)\gamma_P} \cdot x_n(t).$$

The key difference from Kyle (1985) is that the residual demand curve is specified in terms of trader  $n$ 's rate of trading  $x_n(t)$  rather than in terms of the number of shares traded. This specification makes temporary price impact associated with the speed with which traders build their positions economically relevant in a manner not present in the model of Kyle (1985). In the continuous model of Kyle (1985), the informed trader—who buys at rate  $x(t)$ —conjectures that price follows the process  $dP(t) = \lambda \cdot (\sigma_U \cdot dB_U(t) + x(t) \cdot dt)$ , where  $dB_U(t)$  represents random noise trading. If the informed trader buys  $Q$  shares over a period of time  $\Delta t$ , he “walks up the demand schedule” and pays an expected average price

$$(58) \quad P(t) + \frac{1}{2} \lambda \cdot Q.$$

Regardless of how fast he trades, he has no temporary price impact and his permanent price impact is  $\lambda \cdot Q$ .

By contrast, if trader  $n$  gradually buys  $Q$  shares over the time interval  $\Delta t$  in our model, equation (57) implies that the expected average price (e.g., assuming

$H_{-n} = 0$ ) is

$$(59) \quad P(t) + \frac{1}{2} \frac{\gamma_S}{\gamma_P} \frac{1}{N-1} \cdot Q + \frac{1}{(N-1)\gamma_P} \cdot Q/\Delta t.$$

The permanent price impact coefficient  $\gamma_S/\gamma_P \cdot 1/(N-1)$  corresponds to  $\lambda$  in Kyle (1985). The additional term  $1/(N-1) \cdot 1/\gamma_P \cdot Q/\Delta t$  represents temporary price impact; the temporary price impact coefficient is  $1/(N-1)/\gamma_P$ . The temporary price impact cost is proportional to the speed  $Q/\Delta t$  with which trader  $n$  buys  $Q$  shares. It becomes arbitrarily large if he buys that quantity over an arbitrarily short time interval.

Plugging the price impact function (57) into the optimization problem (48), trader  $n$  solves for his optimal consumption and demand schedule. Imperfect competition requires trader  $n$  to take into account both his permanent and temporary price impact in choosing how fast to change his inventory. Trader  $n$  exercises monopoly power in choosing how fast to demand liquidity to profit from innovations in his private information. He also exercises monopoly power in choosing how fast to provide liquidity to the  $N-1$  other traders who, according to trader  $n$ 's beliefs, trade with overconfidence and therefore make supplying liquidity to them profitable.

In equilibrium, the temporary price impact cost parameter  $\gamma_P$  represents compensation to the other traders, arising from adverse selection, for providing trader  $n$  with liquidity quickly. Intuitively, the symmetry of equilibrium trading strategies requires traders to believe they are being adequately compensated for both supplying and demanding liquidity in a manner consistent with market clearing.

### Conjectured Value Function.

We conjecture and verify that the value function  $V(M_n, S_n, D, \hat{H}_n, \hat{H}_{-n})$  has the specific quadratic exponential form

$$(60) \quad V(M_n, S_n, D, \hat{H}_n, \hat{H}_{-n}) = -\exp \left( \psi_0 + \psi_M \cdot M_n + \frac{1}{2} \psi_{SS} \cdot S_n^2 + \psi_{SD} \cdot S_n D + \psi_{Sn} \cdot S_n \hat{H}_n + \psi_{Sx} \cdot S_n \hat{H}_{-n} + \frac{1}{2} \psi_{nn} \cdot \hat{H}_n^2 + \frac{1}{2} \psi_{xx} \cdot \hat{H}_{-n}^2 + \psi_{nx} \cdot \hat{H}_n \hat{H}_{-n} \right).$$

The nine constants  $\psi_0$ ,  $\psi_M$ ,  $\psi_{SS}$ ,  $\psi_{SD}$ ,  $\psi_{Sn}$ ,  $\psi_{Sx}$ ,  $\psi_{nn}$ ,  $\psi_{xx}$ , and  $\psi_{nx}$  have values consistent with a steady state equilibrium.

The term  $\psi_M$  measures the utility value of cash. The terms  $\psi_{SS}$ ,  $\psi_{SD}$ ,  $\psi_{Sn}$ , and  $\psi_{Sx}$  measure the utility value of risky asset holdings. The terms  $\psi_{nn}$ ,  $\psi_{xx}$ , and  $\psi_{nx}$  capture the value of future trading opportunities based on current public and private information. The value of trading on innovations to future information is built into the constant term  $\psi_0$ .

## Characterization of Steady-State Symmetric Equilibrium with Linear Trading Strategies and Quadratic Value Functions.

To solve the trader's optimization problem, we use the “no-regret” approach found in Kyle (1989). Instead of solving for a demand function which depends on price, we suppose instead that the trader observes his equilibrium residual supply schedule, which reveals the value of  $H_{-n}(t)$ , and then picks the optimal point on this residual supply schedule. We show that this optimal point can be implemented with a linear demand schedule.

To solve for a steady state equilibrium, it is necessary simultaneously to determine values for the four  $\gamma$ -parameters defining the optimal demand schedule in equation (55) and the nine  $\psi$ -parameters defining the value function in equation (60). The solution to these equations is discussed in the Appendix. We obtain the following theorem:

**THEOREM 2: *Characterization of Equilibrium.*** *There always exists a no-trade equilibrium (with no well-defined price). In addition, there may exist a steady state equilibrium with symmetric linear flow trading strategies of the form conjectured in equation (55) and a value function  $V(M_n, S_n, D, \hat{H}_n, \hat{H}_{-n})$  for trader  $n$  satisfying the quadratic conjecture in equation (60). Such an equilibrium has the following properties:*

*The parameters  $\psi_{Sx}$ ,  $\gamma_H$ ,  $\gamma_S$ , and  $\gamma_D$  satisfy*

$$(61) \quad \psi_{Sx} = \frac{N-2}{2}\psi_{Sn}, \quad \gamma_H = \frac{N\gamma_P}{2\psi_M}\psi_{Sn}, \quad \gamma_S = -\frac{(N-1)\gamma_P}{\psi_M}\psi_{SS}, \quad \gamma_D = \frac{\gamma_P}{\psi_M}\psi_{SD}.$$

*The parameters  $\psi_M$  and  $\psi_{SD}$  satisfy*

$$(62) \quad \psi_M = -rA, \quad \psi_{SD} = -\frac{rA}{r + \alpha_D},$$

*and  $\psi_0$  satisfies*

$$(63) \quad \psi_0 = 1 - \log(r) + \frac{1}{r} \left( -\rho + \frac{1}{2}(1 + \hat{A}^2)\psi_{nn} + \frac{1}{2} \left( \frac{1}{N-1} + \hat{A}^2 \right) \psi_{xx} + \hat{A}^2 \psi_{nx} \right).$$

*The six constants  $\gamma_P$ ,  $\psi_{SS}$ ,  $\psi_{Sn}$ ,  $\psi_{nn}$ ,  $\psi_{xx}$ , and  $\psi_{nx}$  satisfy the six polynomial equations (111)-(116) in the Appendix. The second order condition, which requires downward sloping demand curves, implies  $\gamma_P > 0$ . Equilibrium also requires  $\psi_{SS} > 0$  and  $\psi_{Sn} < 0$  (implying  $\gamma_S > 0$  and  $\gamma_H > 0$ ).*

*Define the average of traders' expected growth rates  $\bar{G}(t)$  by*

$$(64) \quad \bar{G}(t) := \frac{1}{N} \sum_{n=1}^N G_n(t),$$

and define the constants  $C_L$  and  $C_G$  by

$$(65) \quad C_L := -\frac{\psi_{Sn}}{2\psi_{SS}}, \quad C_G := \frac{\psi_{Sn}}{2\psi_M} \cdot \frac{N(r + \alpha_D)(r + \alpha_G)}{\sigma_G \Omega^{1/2} \cdot (\tau_H^{1/2} + (N-1)\tau_L^{1/2})}.$$

Then, trader  $n$ 's optimal consumption satisfies equation

$$(66) \quad c_n^*(t) = -\frac{1}{A} \cdot \log(\psi_M \cdot V(\cdot)/A).$$

Trader  $n$ 's optimal flow demand schedule  $x_n^*(t)$  makes inventories  $S_n(t)$  a differentiable function of time such that

$$(67) \quad x_n^*(t) = \frac{dS_n(t)}{dt} = \gamma_S \cdot \left( C_L \cdot (\hat{H}_n(t) - \hat{H}_{-n}(t)) - S_n(t) \right).$$

The equilibrium price is

$$(68) \quad P^*(t) = \frac{D(t)}{r + \alpha_D} + \frac{C_G \cdot \bar{G}(t)}{(r + \alpha_D)(r + \alpha_G)}.$$

Mathematical intuition and numerical calculations (as discussed below) suggest that the existence condition for the continuous-time model is the same as the existence condition for the one-period model:

**CONJECTURE 1: *Existence Condition.*** *An equilibrium with trade exists if and only if*

$$(69) \quad \frac{\tau_H^{1/2}}{\tau_L^{1/2}} > 2 + \frac{2}{N-2}.$$

Note there is always a trivial no-trade equilibrium, as in the one-period model. If each trader submits a no-trade demand schedule  $X_n(t, \cdot) \equiv 0$ , then such a no-trade demand schedule is optimal for all traders. This is not a symmetric linear equilibrium in which an auctioneer can establish a meaningful market price.

Equations (67) and (68) imply that the equilibrium with trade has a surprisingly simple structure in which quantities adjust to new information slowly while prices adjust instantaneously. Equation (68) implies that each trader has a target inventory proportional to the difference between his own private signal  $\hat{H}_n(t)$  and the average of other traders' private signals  $\hat{H}_{-n}(t)$  inferred from prices. Each trader continuously moves his inventory towards the target inventory so that the difference decays at rate  $\gamma_S$ . Equation (68) implies that the price is a linear function of the weighted average of all traders' expected growth rates. Price is also the precision-weighted average of the public signal  $H_0(t)$  with precision  $\tau_0^{1/2}$  and the  $N$  private signals  $H_n(t)$  with precision  $[\tau_H^{1/2} + (N-1)\tau_L^{1/2}]/N$  each,

$$(70) \quad P^*(t) = \frac{D(t)}{r + \alpha_D} + \frac{C_G \cdot \sigma_G \cdot \Omega^{1/2}}{(r + \alpha_D)(r + \alpha_G)} \cdot \left( \tau_0^{1/2} \cdot H_0(t) + \frac{\tau_H^{1/2} + (N-1)\tau_L^{1/2}}{N} \cdot \sum_{n=1,\dots,N} H_n(t) \right).$$

The price responds instantaneously to innovations in each trader's private information, so that the average of all signals is immediately revealed. This occurs despite the fact that, to reduce trading costs resulting from adverse selection, each trader intentionally slows down his trading to reduce other traders' estimates of the magnitude of his private signal.

### Implied Price Impact Model.

Since we have an equilibrium model with imperfect competition, we can explicitly calculate the effect on prices if a trader deviates from his optimal inventory policy  $S_n^*(t)$  and instead holds inventories denoted  $S_n(t)$ , assumed to be a differentiable function of time with  $x_n(t) := dS_n(t)/dt$ . As a result of the deviation, the old equilibrium price path  $P^*(t)$  will be changed to a different price path, denoted  $P(t)$ , given by

$$(71) \quad P(t) = P^*(t) + \lambda_S \cdot (S_n(t) - S_n^*(t)) + \lambda_x \cdot (x_n(t) - x_n^*(t)).$$

From equation (57), the constants  $\lambda_S$  and  $\lambda_x$  are given by

$$(72) \quad \lambda_S := \frac{\gamma_S}{(N-1) \cdot \gamma_p}, \quad \lambda_x := \frac{1}{(N-1) \cdot \gamma_p}.$$

The term  $\lambda_S \cdot (S_n(t) - S_n^*(t))$  represents permanent price impact, linear in the numbers of shares. The term  $\lambda_x \cdot (x_n(t) - x_n^*(t))$  represents transitory price impact linear in the rate of trading. Larger trades and faster trading result in larger permanent and temporary price changes.

The concepts of depth, tightness, and resiliency from Black (1971) play out differently in our model than in Kyle (1985). There is no instantaneous depth for "block" trades, i.e., selling 100,000 shares as a block is infinitely expensive. In addition to the permanent impact as in Kyle (1985), there is a temporary impact which increases with the speed of trading; there is a tight spread for traders willing to trade very slowly and significant costs for traders aggressively rushing their orders. For example, selling 100,000 shares over one hour is more expensive than selling 100,000 shares over one day; both executions, however, incur the same permanent impact. Market "liquidity" depends mostly on the "resilience" of prices, which shows up as transitory price impact. The equilibrium links both permanent and transitory impact to deep parameters in the model, such as the mean-reversion of fundamentals and precision of information flow.

### An Existence Condition.

We implement the characterization of equilibrium in Theorem 2 by attempting to solve the equations numerically. As expected, numerical algorithms do not always find an equilibrium with trade satisfying Theorem 2.

Although we have not been able to prove analytically the conditions under which equilibrium exists, extensive numerical experimentation supports the following intuitive argument: Like the one-period model, we expect equilibrium with trade to exist

only if there is enough disagreement. With continuous trading, each trader tries to exercise monopoly power by smoothly walking along the residual demand schedules of other traders rather than by trading a block at one market clearing price. If  $\Delta P$  denotes the price impact of trading some quantity  $\Delta X$  smoothly by walking along a linear residual demand schedule, then the average transactions price incorporates a realized price impact cost of approximately  $\Delta P/2$ . For traders to be willing to take the other side of such smooth trades of their competitors, traders must believe that their competitors' signals are only about "half as precise" (in standard deviation units) as the competitors believe them to be.

To convert this intuitive argument into mathematics, fix all of the exogenous parameters except for the number of "other" traders  $N - 1$  and the "low" precision of their signals  $\tau_L$ . Now allow  $N$  and  $\tau_L$  to vary such that the total precision  $(N - 1)\tau_L$  of other traders is constant. If  $N$  is very large and  $\tau_L$  is very small, there is a huge degree of disagreement, each trader is small relative to the market, and an equilibrium should exist which resembles perfect competition or monopolistic competition. As  $N$  shrinks and  $\tau_L$  increases, eventually a point is reached such that there is not enough disagreement to support an equilibrium. Just before this point is reached, the parameter  $\gamma_P$ —which measures the liquidity of the market—should fall to a value close to zero, the equilibrium should involve very little trade, and the value function should resemble a no-trade equilibrium. The value of  $N$  such that  $\gamma_P = 0$  defines a "critical" value  $N^*$  (not necessarily an integer) such that equilibrium exists if and only if  $N > N^*$ .

This intuitive argument leads to a mathematically precise existence condition derived from the six equations in six unknowns (111)-(116) in the Appendix. Plug  $\gamma_P = 0$  into these equations, representing no market liquidity. Now, holding the other exogenous parameters constant, allow  $N - 1$  and  $\tau_L$  to vary so that  $(N - 1)\tau_L$  is constant and the six equations have a solution. With  $\gamma_P = 0$ , it is clear that  $\psi_{nn} = \psi_{nx} = \psi_{xx} = 0$  solves the last three equations (114)-(116) of the six equations (111)-(116), consistent with the intuition that information has no value if there is no market liquidity. It is straightforward to show that a solution to the first three equations (111)-(113) requires that the critical value  $N^*$  satisfy  $\tau_H/\tau_L = (2+2/(N^*-2))^2$ . This is exactly the same as the existence condition (17) derived in closed form for the one-period model! We therefore conjecture that an equilibrium with trade, consistent with Theorem 2, exists if and only if condition (69) holds.

Our extensive examination of numerical solutions to the six equations (111)-(116) supports this conjecture. We have found that precisely one solution with downward sloping demand schedules ( $\gamma_P > 0$ ) is discovered when existence condition (69) is satisfied. When existence condition (69) is reversed, we find solutions with  $\gamma_P > 0$ , but these solutions imply  $\gamma_S < 0$ , i.e., permanent price impact has the "wrong" sign, which is inconsistent with equilibrium.

## Numerical Comparative Statics Results.

Next, we analyze numerically how the number of traders and the degree of overconfidence affect the equilibrium. We keep the total precision  $\tau$  fixed and vary the degree of overconfidence measured by the ratio of precisions  $\tau_H/\tau_L$  and the degree of competition measured by the number of traders  $N$ .

Figure 1 shows the effect of changes in the degree of overconfidence  $\tau_H/\tau_L$  on parameters  $\gamma_S$ ,  $\gamma_P$ ,  $C_G$ ,  $C_L$ ,  $1/\lambda_S$  and  $1/\lambda_x$ .<sup>6</sup> We change  $\tau_H$  and  $\tau_L$ , keeping the total precision  $\tau$  fixed. Higher values of the ratio  $\tau_H/\tau_L$  correspond to higher degrees of overconfidence. Note that the equilibrium exists only when  $\tau_H$  is sufficiently large relative to  $\tau_L$ , consistent with the existence condition (69); otherwise, the numerical algorithm for solving the system (111)-(116) does not converge to a solution.

The coefficient  $C_G$  decreases monotonically as  $\tau_H/\tau_L$  increases. The more traders disagree with each other, the more they discount actions of others and therefore dampen the equilibrium price sensitivity to the average signal. As the degree of disagreement increases, the rate of inventory adjustment  $\gamma_S$  and the sensitivity of trading rate to prices  $\gamma_P$  increases. The coefficient  $C_L$  (related to target inventories) is a non-monotonic function of  $\tau_H/\tau_L$ . The price impact coefficients  $\lambda_S$  and  $\lambda_x$  decrease, as the degree of disagreement increases, since traders are willing to provide more liquidity to presumably less informed counterparties. As  $\tau_H/\tau_L \rightarrow \infty$ , the model converges to the market of Black (1986), in which each trader thinks that others are noise traders with  $\tau_L = 0$ . In this limit, permanent and temporary market impact  $\lambda_S$  and  $\lambda_x$  converge to zero, inventory adjustment  $\gamma_S$  is infinitely fast, and  $C_G$  converges to some fixed level less than one.

Figure 2 shows the effect of changes in the degree of competition  $N$  on the parameters  $\gamma_S$ ,  $\gamma_P$ ,  $C_G$ ,  $C_L$ ,  $1/\lambda_S$  and  $1/\lambda_x$ . In order to insure that changes in  $N$  do not affect total precision, we assume  $\tau_L = 0$  while holding aggregate precision  $\tau = \tau_H$  fixed.<sup>7</sup> A higher value of  $N$  corresponds to a higher degree of competition. When  $N$  increases, however, the effective risk aversion of the market  $A/N$  becomes small, so prices become consistent with risk neutrality.

Figure 2 shows that the speed of inventory adjustment  $\gamma_S$  increases with the number of traders  $N$ , as each trader thinks that the risk-bearing capacity of the market in aggregate increases, and it becomes less costly for traders to trade aggressively towards their target inventories. The coefficient  $C_L$  defining target inventories increases with  $N$  and converges to an almost constant level after about  $N = 150$ . The coefficient  $C_G$  is monotonically decreasing with  $N$ . The sensitivity of investors' order to market price  $\gamma_P$  increases with  $N$ . Both price impact coefficients  $\lambda_S$  and  $\lambda_x$  decrease as the number of traders  $N$  increases.

<sup>6</sup>We assume that  $\tau = 7.4$ ,  $r = 0.01$ ,  $A = 1$ ,  $\alpha_D = 0.1$ ,  $\alpha_G = 0.02$ ,  $\sigma_D = 0.5$ ,  $\sigma_G = 0.1$ ,  $\tau_0 = \Omega\sigma_G^2/\sigma_D^2 = 0.0054$  and  $N = 100$ .

<sup>7</sup>We assume  $\tau = 1.4$ ,  $r = 0.01$ ,  $A = 1$ ,  $\alpha_D = 0.1$ ,  $\alpha_G = 0.02$ ,  $\sigma_D = 0.5$ ,  $\sigma_G = 0.1$  and  $\tau_0 = \Omega\sigma_G^2/\sigma_D^2 = 0.0279$ .

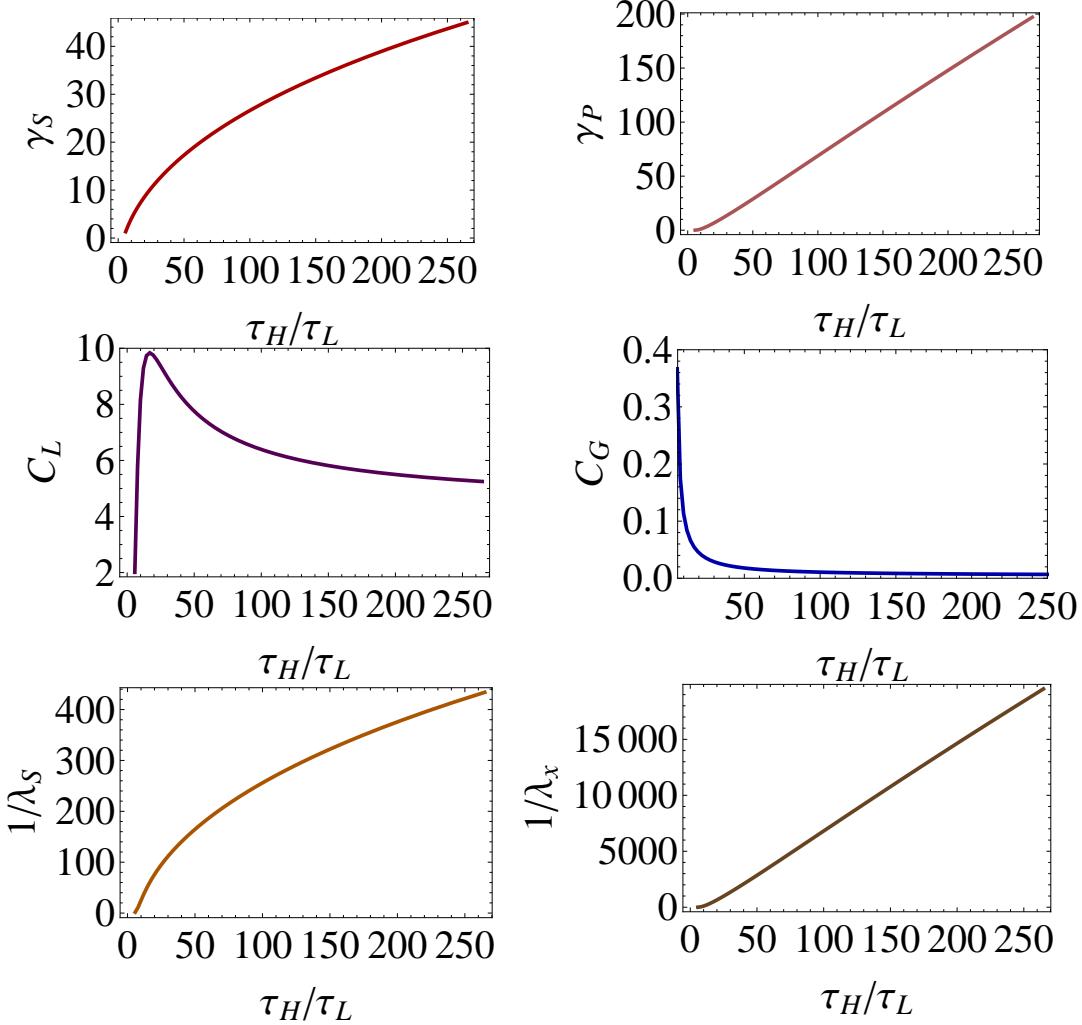


FIGURE 1. COEFFICIENTS  $\gamma_S$ ,  $\gamma_P$ ,  $C_G$ ,  $C_L$ ,  $1/\lambda_S$  AND  $1/\lambda_x$  AGAINST  $\tau_H/\tau_L$  WHILE FIXING  $\tau = 7.4$ .

### III. Implications

This section discusses properties of prices, quantities, and trading strategies in more detail.

#### “Dampened” Prices Reflect a Keynesian Beauty Contest.

Define the “fundamental value” of the risky asset as the expected present value of all future dividends based on all information, discounted at the risk-free rate  $r$ . From the perspective of trader  $n$ , it can be shown that the fundamental value of the risky asset is given by a version of Gordon’s growth formula based on trader  $n$ ’s

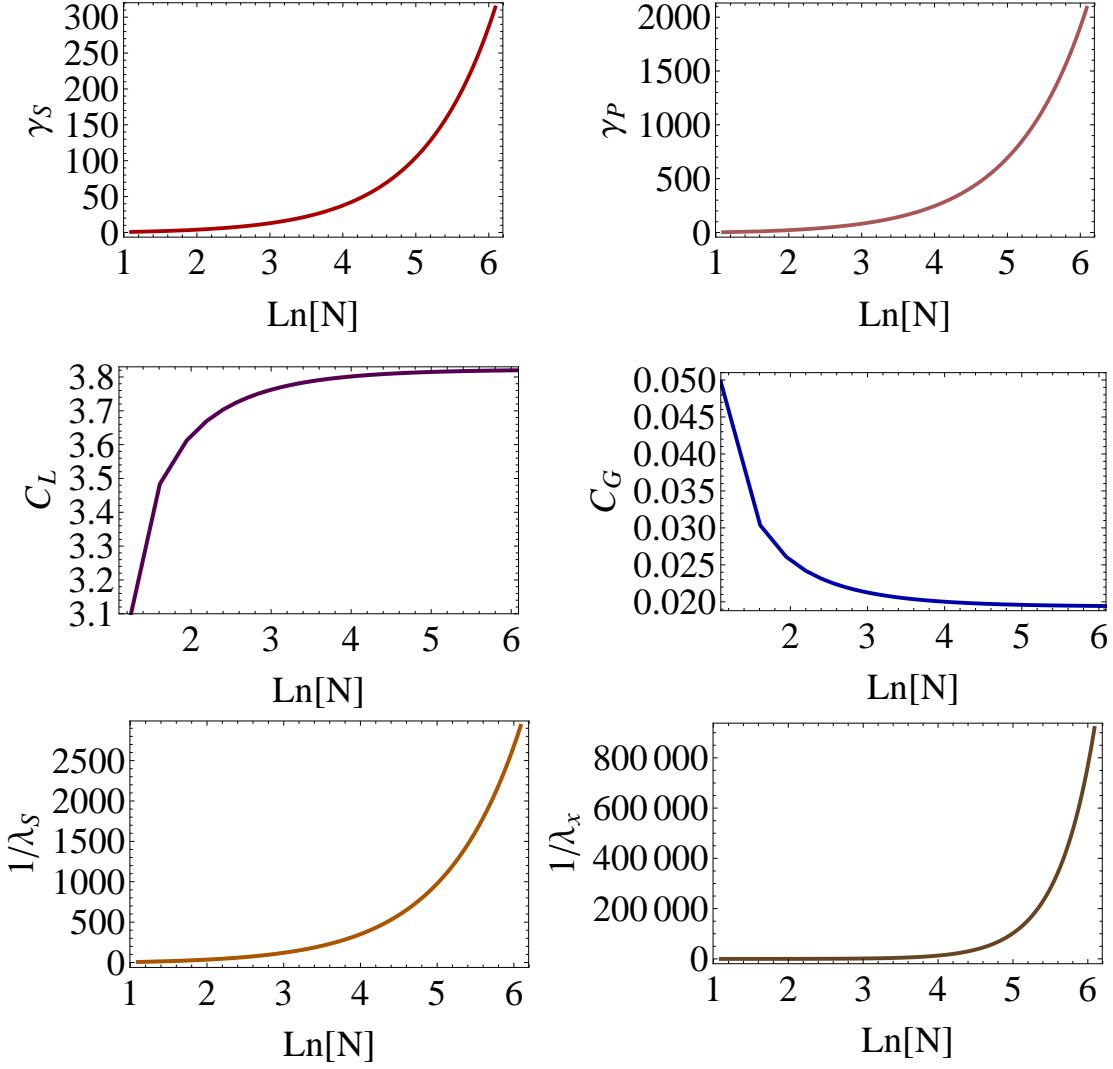


FIGURE 2.  $\gamma_S$ ,  $\gamma_P$ ,  $C_G$ ,  $C_L$ ,  $1/\lambda_S$  AND  $1/\lambda_x$  AGAINST  $N$  WHILE FIXING  $\tau = 1.4$  AND  $\tau_L = 0$

expected growth rate  $G_n(t)$ :

$$(73) \quad F_n(t) = \frac{D(t)}{r + \alpha_D} + \frac{G_n(t)}{(r + \alpha_D)(r + \alpha_G)}.$$

Since the risky asset is in zero net supply, intuition might suggest that the equilibrium price is the average estimate of fundamental value  $\sum_n F_n(t)/N$  obtained by replacing  $G_n(t)$  with  $\bar{G}(t)$  in equation (73). This intuition is precisely consistent with the one-period model. Surprisingly, in the model with continuous trading, this intuition turns out to be wrong! A comparison of equations (68) and (73) reveals that the equilibrium price would be equal to the average of the  $N$  traders' estimates of fundamental value only if  $C_G = 1$ . In our numerical calculations, we always find

$0 < C_G < 1$ ; thus, we conjecture that  $C_G$  must always be less than one. This conjecture implies that, with continuous trading, the equilibrium price is a “dampened” version of Gordon’s growth formula, with dampening factor  $0 < C_G < 1$ . Even if all  $N$  traders unanimously agree on the same expected growth rate  $G_n(t) = \bar{G}(t)$ , the equilibrium price will reflect a dampened implied growth rate  $C_G \cdot \bar{G}(t)$ .

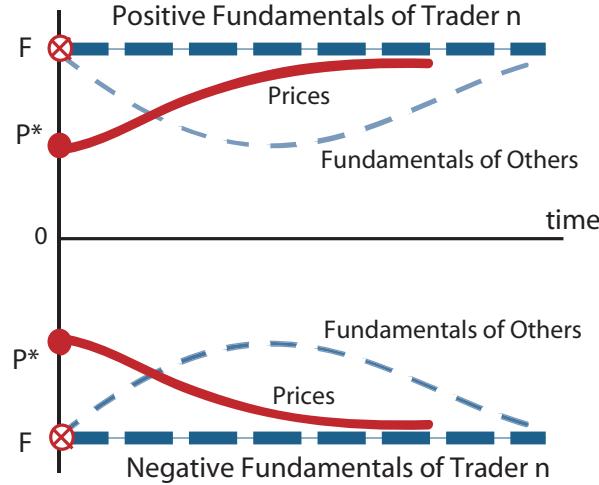


FIGURE 3. DYNAMICS OF PRICES AND FUNDAMENTALS FROM PERSPECTIVE OF A TRADER.

Since the dampening result  $0 < C_G < 1$  contrasts with our one-period model of imperfect competition, the intuitive explanation must be based on having many rounds of trading, not imperfect competition. Indeed, Kyle and Lin (2001) find a similar dampening result in a competitive model of continuous trading.

Figure 3 informally describes the intuition of the dampening effect, which is based on patterns in the evolution of traders’ expectations. The dark dashed (blue) horizontal lines represent the cumulative expected returns trader  $n$  believes he would realize if prices reflected his own expected growth rate  $G_n(t)$ —which he believes to mean-revert to zero at a rate  $\alpha_G$ —and not the average of others’ expected growth rates. Since the line is horizontal, trader  $n$  would be comfortable holding a target inventory of zero, consistent with the market clearing for the zero-net-supply asset. The light dashed (blue) line represents the cumulative expected returns trader  $n$  believes he would realize if prices reflected the average of all traders’ beliefs about expected growth rates and all  $N$  traders started with the same expected growth rate  $G_m(t) = G_n(t) = \bar{G}(t)$ . This line first moves towards zero and then moves back to its initial level, consistent with the interpretation that trader  $n$  believes that the other traders’ estimates of expected growth rates will first mean-revert to zero at a rate faster than  $\alpha_G$  (since other traders’ estimates of the growth rate are based on signals with less precision than those traders thought they had) and then move back towards

the initial value (since trader  $n$  believes his own estimate will be proven correct in the long run). Since all traders expect prices to deviate from the long term mean in the short run, traders will not want to hold inventories of zero, even if prices reflect consensus fundamentals. Traders will want to sell if the consensus growth estimate is positive (top of figure) and buy if the consensus growth estimate is negative (bottom of figure). This leads to an equilibrium price dampened relative to fundamentals, as depicted by the solid (red) lines. In equilibrium, when traders have positive expected growth rates, they may expect returns to be slightly negative in the short run but positive in the long run, eventually reflecting the traders' common expected growth rates  $\bar{G}(t)$ . This intuition supports the conjecture  $0 < C_G < 1$ .

Our model captures the intuition of the beauty contest described by Keynes (1936):<sup>8</sup>

“For most of these persons are, in fact, largely concerned, not with making superior long-term forecasts of the probable yield on an investment over its whole life, but with foreseeing changes in the conventional basis of valuation a short time ahead of the general public. They are concerned not with what an investment is really worth to a man who buys it ‘for keeps,’ but with what the market will value it at, under the influence of mass psychology, three months or a year hence.”

As in Keynes (1936), traders in our model trade based on short-term price dynamics rather than hold-to-maturity values. As Keynes puts it, “it is not sensible to pay 25 for an investment of which you believe the prospective yield to justify value of 30, if you also believe that the market will value it at 20 three months hence.”

Keynes also believed that since financial markets are dominated by short-term speculation rather than long-term enterprise, they are not too different from a casino and exhibit excessive volatility. In contrast to Keynes, short-term trading dynamics dampens price volatility in our model relative to the volatility of fundamental value.

To summarize, each trader believes that equilibrium prices usually differ from fundamentals, prices do not follow a martingale, and price changes are predictable.

<sup>8</sup>“...Professional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole; so that each competitor has to pick, not those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors, all of whom are looking at the problem from the same point of view. It is not a case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practise the fourth, fifth and higher degrees.” Our model implicitly assumes that all traders anticipate the expectations of other traders for arbitrarily higher degrees. Han and Kyle (2013) examine a one-period model in which, instead of agreeing to disagree, traders disagree about higher order beliefs.

### Trading Strategies Follow A Partial Adjustment Process.

With continuous trading, the equilibrium trading strategies have a simple form similar to the one-period model. Let  $S_n^{TI}(t)$  denote the “target inventory” of trader  $n$ , defined as the inventory level  $S_n^{TI}(t)$  such that the informed trader chooses not to trade  $x_n(t) = 0$ . From equation (67), the target inventory  $S_n^{TI}(t)$  is given by

$$(74) \quad S_n^{TI}(t) = C_L \cdot \left( \hat{H}_n(t) - \hat{H}_{-n}(t) \right).$$

Trader  $n$  targets a long position if his own signal  $H_n(t)$  is greater than the average signal of other traders  $H_{-n}(t)$  and targets a short position if his own signal is less than the average signal of others. Trader  $n$  follows a partial adjustment strategy, with his inventory  $S_n(t)$  converging towards its optimal level  $S_n^{TI}(t)$  at rate  $\gamma_S$ :

$$(75) \quad x_n(t) = \frac{dS_n(t)}{dt} = \gamma_S \cdot (S_n^{TI}(t) - S_n(t)).$$

While the target inventory levels  $S_n^{TI}(t)$  and the speed of trading  $x_n(t)$  change like a diffusion of order  $dt^{1/2}$  due to arrival of new information, actual inventories change differentiably at rate of order  $dt$ .

When a trader observes a new signal, he updates his estimate of the growth rate, recalculates his target inventory, and immediately adjusts the rate of trading towards the new target. Since block trades are infinitely expensive, a trader does not trade immediately to the new target but rather adjusts his inventories slowly, taking into account his market power. As soon as a trader changes the speed of trading, however, the price of a risky asset instantaneously moves to a new equilibrium level, even though a trader has not traded a single share yet.

Equations (74) and (75) imply that inventories  $S_n(t + s)$  have the integral representation

$$(76) \quad S_n(t + s) = e^{-\gamma_S \cdot s} \cdot \left( S_n(t) + \int_{u=t}^{t+s} e^{-\gamma_S \cdot (t-u)} \cdot \gamma_S \cdot C_L \cdot (\hat{H}_n(u) - \hat{H}_{-n}(u)) \cdot du \right).$$

The equation has a simple intuition. Traders accumulate inventories gradually based on their current level of disagreement  $\hat{H}_n(t) - \hat{H}_{-n}(t)$ ; inventories accumulated based on past disagreement are gradually liquidated at rate  $\gamma_S$ . If signals  $\hat{H}_n(t)$  and  $\hat{H}_{-n}(t)$  do not change, the price does not change, but trader  $n$  will continue to trade based on the level of his “past” disagreement with the market  $\hat{H}_n(t) - \hat{H}_{-n}(t)$ .

### Transaction Costs Depend on Quantities and Speed of Trading.

Equation (71) can also be used to calculate the out-of-equilibrium price effect of a “new” trader  $n = N + 1$  who silently enters the market and acquires inventories  $\bar{S}_{N+1}(t)$  at a rate  $\bar{x}_{N+1}(t)$ , unbeknownst to the other  $N$  traders. Since the new trader does not actually trade in equilibrium, we assume  $S_{N+1}(t) = x_{N+1}(t) = 0$  in equation (71). By affecting prices, the new trader incurs permanent and temporary

price impact costs, denoted  $\tilde{C}$ . We measure these costs using the concept of implementation shortfall, as described by Perold (1988). The expected price impact costs are given by

$$(77) \quad E\{\tilde{C}\} = E \left\{ \int_{u=t}^{\infty} (P(u) - P^*(u)) \cdot \bar{x}(u) \cdot du \right\}.$$

The actual expected implementation shortfall depends on how the new trader trades. Here are two simple examples.

*Example 1:* Suppose the new trader  $N + 1$  enters the market at date  $t = 0$  and acquires a random block of shares  $\tilde{B}$ , uncorrelated with signals  $H_n(t)$ ,  $n = 1, \dots, N$ , by trading at the constant rate  $\bar{x}(t) = \tilde{B}/T$  over the interval  $[0, T]$ . Using  $\lambda_S = \gamma_S \cdot \lambda_x$  from equation (72), new trader  $N + 1$ 's expected “implementation shortfall” is given by

$$(78) \quad E\{\tilde{C}\} = \left( \lambda_S + \frac{\lambda_x}{T/2} \right) \cdot \frac{\tilde{B}^2}{2} = \lambda_S \cdot \left( 1 + \frac{1}{\gamma_S} \cdot \frac{1}{T/2} \right) \cdot \frac{\tilde{B}^2}{2}.$$

*Example 2:* Suppose instead that the new trader enters the market at date  $t = 0$  and acquires the random inventory  $\tilde{B}$  by trading at rate  $x_{N+1}(t) = \gamma \cdot (\tilde{B} - \bar{S}_{N+1}(t))$ . Then his inventory evolves as  $\bar{S}(t) = \tilde{B} \cdot (1 - e^{-\gamma t})$ , with  $\bar{S}(t) \rightarrow \tilde{B}$  as  $t \rightarrow \infty$ . The implementation shortfall is given by

$$(79) \quad E\{\tilde{C}\} = \left( \lambda_S + \gamma \cdot \lambda_x \right) \cdot \frac{\tilde{B}^2}{2} = \lambda_S \cdot \left( 1 + \frac{\gamma}{\gamma_S} \right) \cdot \frac{\tilde{B}^2}{2}.$$

If the new trader chooses a speed of execution  $\gamma$  equal to the equilibrium speed of execution  $\gamma_S$ , then the permanent cost  $\lambda_S \cdot \tilde{B}^2/2$  is equal to the temporary cost  $\lambda_S \cdot \tilde{B}^2/2 \cdot \gamma/\gamma_S$ ; both are equal to one half of the total cost  $E\{C\} = \lambda_S \cdot \tilde{B}^2$ . The intuition for this result is the same as the intuition for traders “in the model.” Since each trader is trying to walk up the demand schedule of the other traders by slowing down his trading, all traders cannot trade simultaneously as price discriminating monopolists. In a symmetric equilibrium, each trader expects to break even providing liquidity to other traders. This requires each trader to pay out his potential monopoly profit from walking the demand curve,  $\lambda_S \cdot \tilde{B}^2/2$  (exactly half the costs incurred due to permanent impact), to others in the form of temporary price impact. Note that each trader still expects to make a profit because traders disagree with one another about the fundamental value of the asset.

In both examples, faster execution leads to larger temporary price impact but has no effect on permanent price impact. Infinitely fast block trades with  $T \rightarrow 0$  or  $\gamma \rightarrow \infty$  become infinitely expensive due to temporary price impact costs. Infinitely slow trades with  $T \rightarrow \infty$  or  $\gamma \rightarrow 0$ —in which the new trader slowly walks up or down

the demand schedule like a price discriminating monopolist—incur only a permanent impact cost  $\lambda_S \cdot \tilde{B}^2/2$  and no temporary impact costs. If the new trader unwinds his position at a later date, the new trader recovers the permanent impact costs but incurs the temporary impact costs once again.

Almgren and Chriss (2000) use a functional form for price impact like equation (71) because of its computational simplicity. We show that this functional form arises endogenously in a model of informed trading. Obizhaeva and Wang (2013) suggest an alternative model in which—rather than decaying instantaneously when a trader stops trading—temporary price impact decays gradually at an exponential rate. This approach is similar in spirit to Kyle (1985), in which all of the price impact of the informed trader is permanent, but all of the price impact of noise trades dies out linearly over time between the time the noise trades are made and the end of trading. The profits from these trades allow market makers to earn “spread profits” from noise traders large enough to cover their losses trading against the informed trader. If noise traders smoothed their trading out over a short period of time, walking the demand schedule of the market makers, the average temporary price impact cost would be cut in half to  $\lambda \cdot dB^2/2$ . As a result, market making would be unprofitable for any level of market depth  $\lambda$ . Intuitively, the temporary price impact costs in our model are necessary to make it profitable for traders to provide liquidity to one another when all traders smooth their trading to economize on temporary price impact costs.

### Fast Execution Can Lead To Flash Crashes.

Equation (71) can also be used to describe what happens if trader  $n$  “in the model,”  $1 \leq n \leq N$ , silently deviates from his optimal trading strategy. Suppose that at date  $T$ , instead of trading towards his target inventory at equilibrium rate  $\gamma_S$ , trader  $n$  deviates from his optimal strategy by trading towards his target inventory at some arbitrarily faster or slower rate  $\gamma \neq \gamma_S$  and implements the strategy

$$(80) \quad \bar{x}_n(T+t) = \gamma \cdot (S_n^{TI}(T+t) - \bar{S}_n(T+t)), \quad t \geq 0.$$

Equation (80) becomes the equilibrium strategy when  $\gamma = \gamma_s$ . The inventory level  $\bar{S}_n(t)$  then coincides with the equilibrium inventory level  $S_n(t)$  before date  $T$ , but deviates afterwards, so that for any  $t \geq 0$ ,

$$(81) \quad \bar{S}_n(T+t) = e^{-\gamma \cdot t} \cdot \left( S_n(T) + \int_{u=T}^{T+t} e^{-\gamma \cdot (T-u)} \cdot \gamma \cdot C_L \cdot (\hat{H}_n(u) - \hat{H}_{-n}(u)) \cdot du \right).$$

To consider an analytically tractable example, suppose that just before date  $T$ , we have  $\hat{H}_n(T^-) > 0$  and  $\hat{H}_{-n}(T^-) = 0$ ; furthermore, suppose trader  $n$  happens to hold his positive target inventory, which is positive since  $\hat{H}_n(T^-) > 0$ :

$$(82) \quad S_n(T) = S_n^{TI}(T^-) = C_L \cdot \hat{H}_n(T^-) > 0.$$

Equation (57) implies

$$(83) \quad P(T^-) = \frac{\gamma_S}{(N-1)\gamma_P} \cdot S_n(T) > 0.$$

Now suppose that the information suddenly changes at date  $T$  so that the value of trader  $n$ 's  $\hat{H}_n(t)$  drops from a positive number to  $\hat{H}_n(T) = 0$ . At the same time, suppose signal  $\hat{H}_{-n}(t)$  remains equal to zero. As a result, trader  $n$  begins to liquidate his inventory, moving from inventory  $S_n(T) > 0$  towards a new target inventory of zero. Using equation (81), trader  $n$ 's expected inventories at dates  $T + t$ ,  $t > 0$ , conditional on information at date  $T$ , are given by

$$(84) \quad E_T^n\{\bar{S}_n(T+t)\} = e^{-\gamma t} \cdot S_n(T).$$

Since  $E_T^n\{P(T+t)\} = 0$  for equilibrium expected prices  $P(t)$ , equation (71) implies that expected out-of-equilibrium prices  $\bar{P}(t)$  are given by

$$(85) \quad E_T^n\{\bar{P}(T+t)\} = -\frac{\gamma - \gamma_S}{(N-1)\gamma_P} \cdot e^{-\gamma t} \cdot S_n(T).$$

If trader  $n$ , like other traders, follows the equilibrium strategies with  $\gamma = \gamma_S$ , then the price immediately falls to zero and is expected to stay there. Otherwise interesting price patterns arise.

Figure 4 shows expected paths of future prices based on equation (85) in panel A and future inventories based on equation (84) in panel B for different horizons, as trader  $n$  sells  $S_n(T)$  shares over time.

We consider the two cases: Trader  $n$  implements his execution (1) at a rate five times faster than the equilibrium rate ( $\gamma = \gamma_S \cdot 5$ , dotted lines) and (2) at a rate five times slower than the equilibrium rate ( $\gamma = \gamma_S/5$ , solid lines). Since price dynamics depend on the total precision  $\tau$ , we compare a model with lower precision and slower mean reversion (dark color) to a model with higher precision and faster mean reversion (light color).<sup>9</sup>

When a trader sells at a rate five times slower than the equilibrium rate,  $\gamma = \gamma_S/5$ , the price immediately drops only about 1/5 as much as in equilibrium. Each trader believes that the higher profits on the early trades at initially better prices are more than offset by lower profits on later trades resulting from information being incorporated into prices by the trading of others before the later trades are completed.

When a trader sells at a rate five times faster than the equilibrium rate,  $\gamma = \gamma_S \cdot 5$ , the price is expected to drop sharply, by about five times as much as in equilibrium.

<sup>9</sup>We assume  $S_n(T) = 1,000$  shares. For the higher precision model, we assume  $\tau = 14.09$  with  $\tau_0 = 0.0028$ ,  $\tau_L = 0.0708$  and  $\tau_H = 7.08$ , implying equilibrium price  $P_1(T^-) = 2.045$  and equilibrium  $\gamma_{S2} = 50.6$  (light color). For the lower precision model, we assume  $\tau = 9.95$  with  $\tau_0 = 0.004$ ,  $\tau_L = 0.05$  and  $\tau_H = 5.00$ , implying equilibrium price  $P_2(T^-) = 2.896$  and equilibrium  $\gamma_{S1} = 35.8$  (dark color). In both models, the other exogenous parameter assumptions are  $r = 0.01$ ,  $A = 1$ ,  $\alpha_D = 0.1$ ,  $\alpha_G = 0.02$ ,  $\sigma_D = 0.5$ ,  $\sigma_G = 0.1$ , and  $N = 100$ ,  $D(T) = 0$ .

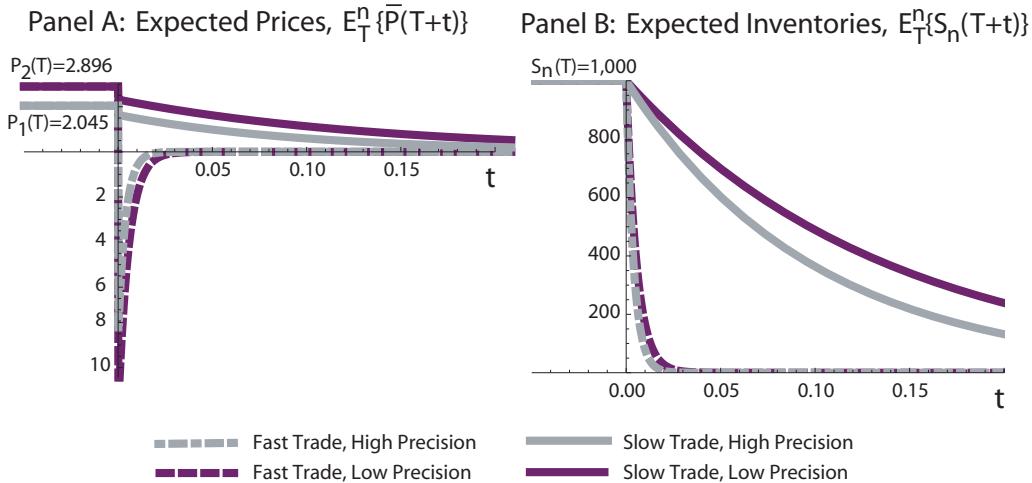


FIGURE 4. THE EXPECTED PRICE AND INVENTORIES DYNAMICS.

Speeding up execution exacerbates transitory price impact and elevates transactions costs. As the price converges to the equilibrium, the price path exhibits a distinct V-shaped pattern. Figure 4 shows that prices converge to their long-term equilibrium values more quickly in the high precision model than in the low precision model.

This calibrated price response is very similar to the price patterns observed during the flash crash of May 6, 2010, when the E-mini S&P 500 futures price plunged by 5% over a five-minute period and then quickly recovered all of the earlier losses after the CME's pre-programmed circuit breakers triggered a five-second pause in futures trading. Staffs of the CFTC and SEC (2010a,b) reported that the flash crash was triggered by an automated execution algorithm that sold S&P 500 E-mini futures worth approximately \$4 billion. Kyle and Obizhaeva (2013) note that market microstructure invariance would imply a price impact of only 0.61% and attributed the difference from actual price changes to unusually fast execution of the order. Indeed, the entire order was executed over a twenty-minute period, while orders of similar magnitude would normally be expected to be executed over horizons of at least several hours. Our model implies that the selling should occur *after* prices crash, while the market recovers. This is reasonably consistent with the pattern observed during the flash crash, since most of the \$4 billion in selling took place after the market had crashed and while prices were recovering.

Trading faster or slower than the equilibrium rate will also affect price volatility. For example, with 100 traders, a trader who trades five times faster than the equilibrium rate will contribute 25 times more to returns variance, initially increasing returns variance by about 25 percent.

### The Value Function and Marking to Market.

The value function of each trader is such that the implied valuation of his risky position reflects its illiquidity as well as the potential disagreement between a trader and the rest of the market about its fundamental value.

In trader  $n$ 's value function (60), the value of inventories  $S_n(t)$  can be expressed in monetary units by scaling by  $\psi_M$ . Letting  $P_n(t)$  denote the dollar value of one unit of inventories in trader  $n$ 's value function, we have

$$(86) \quad P_n(t) = \frac{\psi_{SD}}{\psi_M} \cdot D(t) + \frac{\psi_{Sn}}{\psi_M} \cdot \hat{H}_n(t) + \frac{\psi_{Sx}}{\psi_M} \cdot \hat{H}_{-n}(t) + \frac{\psi_{SS}}{2\psi_M} \cdot S_n(t).$$

The  $\psi_{SS}$  term adjusts the cost of carrying inventories. The  $\psi_{SD}$  term, with  $\psi_{SD}/\psi_M = 1/(r+\alpha_D)$  from equation (62), measures the cash-flow value of dividends. Intuitively, the  $\psi_{Sn}$  and  $\psi_{Sx}$  terms average together in some manner— influenced by the level of market liquidity— both trader  $n$ 's expectation of the value of cash flows and the market price  $P(t)$ .

If the market has almost no liquidity, the trader's value of inventories  $P_n(t)$  will be close to a no-trade value reflecting only trader  $n$ 's estimate of the expected present value of cash flows, discounted for risk. Using equations (37) and (73), the expected present value of cash flows is given by

$$(87) \quad F_n(t) = \frac{1}{r + \alpha_D} \cdot D(t) + \frac{\sigma_G \Omega^{1/2} \cdot \tau_H^{1/2}}{(r + \alpha_D)(r + \alpha_G)} \cdot \hat{H}_n(t) + \frac{\sigma_G \Omega^{1/2} \cdot (N - 1) \tau_L^{1/2}}{(r + \alpha_D)(r + \alpha_G)} \cdot \hat{H}_{-n}(t).$$

If the market is almost perfectly competitive, implying that trader  $n$  can convert his inventories into cash at current market prices with no market impact costs, then the value of inventories  $P_n(t)$  will correspond closely to the mark-to-market value  $P(t)$ , which is given using equation (57) by

$$(88) \quad P(t) = \frac{\psi_{SD}}{\psi_M} \cdot D(t) + \frac{\psi_{Sn}}{\psi_M} \cdot \frac{1}{2} \cdot \hat{H}_n(t) + \frac{\psi_{Sn}}{\psi_M} \cdot \frac{N - 1}{2} \cdot \hat{H}_{-n}(t).$$

Using  $\psi_{Sx} = (N - 2)\psi_{Sn}/2$  from equation (61), it can be shown from comparison of equations (86) and (88) that

$$(89) \quad P_n(t) = P(t) + \frac{\psi_{Sn}}{\psi_M} \cdot \frac{1}{2} \cdot (\hat{H}_n(t) - \hat{H}_{-n}(t)) + \frac{1}{2} \cdot \frac{\psi_{SS}}{\psi_M} \cdot S_n(t).$$

The difference between trader  $n$ 's private valuation  $P_n(t)$  and the market price  $P(t)$  has an interesting economic interpretation. Using equation  $S_n^{TI}(t) = C_L \cdot (\hat{H}_n(t) - \hat{H}_{-n}(t))$  and definitions of parameters  $\psi_{Sn}$ ,  $\psi_{SS}$ ,  $\gamma_S$ , and  $\lambda_S$  from equations (61), (65), (72), it can be shown that

$$(90) \quad P_n(t) = P(t) + \lambda_S \cdot S_n^{TI}(t) - \frac{1}{2} \lambda_S \cdot S_n(t).$$

In privately valuing his risky inventories at  $P_n(t)$ , trader  $n$  effectively makes two adjustments to the market price  $P(t)$ . First, he marks the market price upwards by  $\lambda_S \cdot S_n^{TI}(t)$ , the discrepancy between the trader's appraised value and the market price, which is determined by the intuitive optimality condition: The benefits of acquiring each share at the market price are greater than the *marginal* costs of future liquidation of his target inventories gradually over time. Second, he also marks the market price downwards by  $\frac{1}{2}\lambda_S \cdot S_n(t)$ , the *average* permanent impact cost that he would incur if he liquidated his current inventories  $S_n(t)$  at the slowest rate possible. For example, if the market price is \$20 per share, but the trader thinks that the asset is worth \$40 per share and expects that the average cost of liquidating the position will be \$10 per share, then he assigns the value of \$30 per share of the asset in his value function.

From trader  $n$ 's value function (60), the dollar value of future trading opportunities based on current information is given by

$$(91) \quad \frac{1}{2} \cdot \frac{\psi_{nn}}{\psi_M} \cdot \hat{H}_n(t)^2 + \frac{1}{2} \cdot \frac{\psi_{xx}}{\psi_M} \cdot \hat{H}_{-n}(t)^2 + \frac{\psi_{nx}}{\psi_M} \cdot \hat{H}_n(t) \cdot \hat{H}_{-n}(t)$$

Intuition suggests that the value of trading opportunities should always be non-negative: a trader will not trade on information unless the trader expects the trades to be profitable. Consistent with this intuition, our numerical results always find that (91) is a positive definite quadratic form. The symmetry of this quadratic form about zero is consistent with the symmetry of the zero-net-supply model about zero; long positions are just as profitable as short positions.

### Market Liquidity.

Practitioners often observe that market liquidity is ephemeral. The mathematics of our model suggests related hypotheses concerning the stability or continuity of liquidity.

The four equations (105) result from imposing symmetry on symmetric linear strategies. While intuition suggests that these equations might determine the four  $\gamma$ -parameters  $\gamma_D, \gamma_H, \gamma_S$  and  $\gamma_P$  as functions of the nine  $\psi$ -parameters, these equations do not actually determine  $\gamma_P$  as a function of the  $\psi$ -parameters; instead, they determine the three  $\gamma$ -parameters  $\gamma_D, \gamma_H$ , and  $\gamma_S$  and generate the  $\psi$ -parameter restriction  $\psi_{Sx} = (N/2 - 1)\psi_{Sn}$ . The sensitivity of trading to price  $\gamma_P$  is left undetermined.

The intuition for this result is that the objective function is a quadratic function of  $x_n(t)/((N - 1)\gamma_P)$ . The first order condition determines only this ratio, i.e., the optimal strategy  $x_n(t)$  for each level of market liquidity  $(N - 1)\gamma_P$ : Traders choose more aggressively when market depth is high. The implied value function turns out to be linear in the level of market liquidity  $(N - 1)\gamma_P$ .

Mathematically, the steady-state level of market liquidity  $(N - 1)\gamma_P$  must be such that the steady-state values of the  $\psi$ -parameters in the value function remain constant over time at just the right ratio consistent with equilibrium. If the equilibrium

were not in a steady state, the six equations (111)-(116) in the Appendix would not be polynomials but would instead be differential equations with the zeros on the right-hand-side changed to derivatives of the corresponding  $\psi$ -parameters with respect to time. If market liquidity  $(N - 1)\gamma_P$  is different from its equilibrium level, the values of  $\psi$ -parameters such as  $\psi_{Sn}$  and  $\psi_{Sx}$  will over time wander away from their equilibrium levels. Indeed, the modified equations (112) and (113) will imply that  $\psi_{Sn}$  and  $\psi_{Sx}$  will change by the same margin proportional to  $\gamma_P(N - 1)\psi_{Sn}\psi_{SS}/(rA)$  but in the opposite directions, and the equality  $\psi_{Sx} = (N/2 - 1)\psi_{Sn}$  will eventually be violated. The economic intuition for this is the following: If the equality  $\psi_{Sx} = (N/2 - 1)\psi_{Sn}$  becomes an inequality, then (depending on the direction of the inequality) *either* traders will want to supply liquidity more aggressively than they demand it *or* traders will want to demand liquidity more aggressively than they supply it—for any level of market liquidity  $(N - 1)\gamma_P$ .

Thus, the level of market liquidity  $(N - 1)\gamma_P$  is pinned down at a value consistent with the steady-state dynamics in equations (111)-(116) and the equilibrium  $\psi$ -parameter restriction. At the same time, traders choose an intensity of trading which is consistent with the equilibrium market depth.<sup>10</sup>

When we think about how this delicate relationship between the level of market liquidity and desire to supply and demand liquidity plays out in actual markets, we are left with the intuition that the level of liquidity is probably somewhat indeterminate over very short period of time. The model says literally that private information is manufactured at a constant rate over time. As a practical matter, more private information might be manufactured on trading days than weekends, more during the day than overnight. The spirit of the model is that, at any time during the week, market liquidity is in some sense proportional to the rate at which private information is being manufactured. But what happens if this proportionality is violated? For example, what happens if private information is manufactured at a constant rate twenty-four hours per day, seven days per week, while market liquidity is much greater during business hours than in the evenings? In this case, the equality  $\psi_{Sx} = (N/2 - 1)\psi_{Sn}$  will be violated, but perhaps in an economically insignificant manner. Traders will be just about as happy with constant market liquidity twenty-four hours per day and three times higher liquidity during eight business hours with closed markets the rest of the day. Whether this intuition is correct is an interesting

<sup>10</sup>A somewhat analogous delicacy in determination of equilibrium arises in the continuous model of Kyle (1985), where the optimization problem of the informed trader is linear in the intensity with which the informed trader trades. This linearity does not place a restriction on the trading strategy of the informed trader by itself, but instead requires market depth to be constant. If market depth were not constant, then either the informed trader would try to destabilize prices and generate unbounded profits or he would want to incorporate information into prices aggressively. The informed trader's optimal trading strategy must be consistent with a constant level of market depth, but this consistency condition comes from the conditions determining market depth, not the conditions determining the optimal trading strategy. In equilibrium, the informed trader ultimately does not care how aggressively he incorporates his private information into prices. Nevertheless, to sustain the equilibrium, the informed trader must choose to trade with an intensity that incorporates information into prices at a constant rate, consistent with constant market depth.

issue for further research.

## IV. Conclusion

We have described a steady-state model of continuous trading, in which trading reflects both overconfidence and market power. This model provides a framework for thinking about how the dynamics of trading affect market liquidity, transaction costs, and market prices.

The idea that securities markets offer a flow equilibrium rather than a stock equilibrium may seem far-fetched at first glance. Yet recent trends in the way liquidity is supplied and demanded in electronic markets are in many ways consistent with the way our model predicts liquidity to be supplied and demanded.

Relative to order processing by human clerks, electronic processing of orders has dramatically reduced the fixed costs of executing an order. The result has been a dramatic reduction in the average size of the trades “printed” in price reporting systems and a correspondingly large increase in the number of orders and messages routed to various trading venues. In previous decades, a trader may have purchased 100,000 shares of stock in a single block trade for 100,000 shares. Nowadays, a trader might execute 1,000 orders for 100 shares each, over a matter of several hours. Although the trader’s inventory would not theoretically be a differentiable function of time, there is a small economic difference between purchasing continuously at the constant rate of 100 shares per minute and an “order shredding” strategy of purchasing numerous 100-share increments at random (poisson) time intervals spaced about one minute apart.

Our model predicts that there should be vanishingly small market depth available at a given point in time; instead, market depth is predicted to be made available only over time. In today’s markets, the actual level of market depth available at the “top of the book”—i.e., at the best bid and the best offer—is influenced by tick size (the smallest units in which prices are allowed to fluctuate) and by rules for time and price priority. Since time priority mandates execution of the older resting limit orders before newer ones with the same limit price, time priority encourages traders to place bids and offers into the limit order book in order to have their orders executed ahead of others who want to trade at the same price at around the same time. Relative to our model, time priority creates an externality among traders which results in more depth in the limit order book than would otherwise be present. This microstructure externality may be more important when minimum tick size is large. Between 1997 and 2001, the minimum tick size in the U.S. equity markets was reduced by a factor of 12.5 from one-eighth of a dollar to one cent. Our model implies an infinitesimal tick size. Thus, today’s markets probably have more instantaneous market depth available than our theory would imply, but they may have less instantaneous depth available than they would have if tick size were larger.

Our model of smooth order flow implements ideas about market liquidity described informally by Black (1995). Black envisioned a future frictionless market for exchanges as “an equilibrium in which traders use indexed limit orders at different

levels of urgency but do not use market orders or conventional limit orders.” In that equilibrium, there is no conventional liquidity available for market orders and conventional limit orders. Placement of indexed orders onto the market moves the price by an amount increasing in level of urgency. Our model implements Black’s intuition in a precise mathematical manner.

Algorithms for executing orders to trade stocks and futures contracts have for years been incorporating the idea of urgency. For example, algorithms based on VWAP (“Volume Weighted Average Price”) have become popular. In a VWAP trade, a trader chooses a target number of shares to trade, a time frame (say one day) and a participation rate (say 5% of volume). The higher the participation rate, the greater the trader’s impatience.

In the future, exchanges might change order matching rules to implement limit orders conforming to the intuition of our model. For example, exchanges might approximate our flow model by having frequent batch auctions, say once per second, consistent with Budish, Cramton and Shim (2013). Limit orders could easily be implemented with a time parameter. For example, a trader who might in today’s market place a limit order to buy 10,000 shares at a price of \$40 per share might instead enter an order to purchase one share per second at a price of \$40 or better. The order would be filled over 10,000 seconds if the price moved below \$40 per share and stayed there. It would take longer, or perhaps never be fully executed, if the price moved above \$40 per share and stayed there. Clearly, such orders would reduce the high levels of message traffic which would be generated by similar strategies using thousands of conventional limit orders and order cancellations.

## REFERENCES

- Allen, Franklin, Stephen Morris, and Hyun Song Shin.** 2006. “Beauty Contests and Iterated Expectations in Asset Markets.” *Review of Financial Studies*, 19(3): 719–752.
- Almgren, Robert, and Neil Chriss.** 2000. “Optimal Execution of Portfolio Transactions.” *Journal of Risk*, 3: 5–39.
- Almgren, Robert, Chee Thum, Emmanuel Hauptmann, and Hong Li.** 2005. “Direct Estimation of Equity Market Impact.” *Risk*, 18(7): 58–62.
- Banerjee, Snehal, Ron Kaniel, and Ilan Kremer.** 2009. “Price Drift as an Outcome of Differences in Higher-Order Beliefs.” *Review of Financial Studies*, 22(9): 3707–3734.
- Black, Fischer.** 1971. “Towards a Fully Automated Exchange, Part I.” *Financial Analysts Journal*, 27: 29–34.
- Black, Fischer.** 1986. “Noise.” *The Journal of Finance*, 41: 529–543.
- Black, Fischer.** 1995. “Equilibrium Exchanges.” *Financial Analysts Journal*, 51: 23–29.

- Black, Fischer, and Jack L. Treynor.** 1973. “How to Use Security Analysis to Improve Portfolio Selection.” *Journal of Business*, 46(1): 66–86.
- Brunnermeier, Markus K., and Lasse H. Pedersen.** 2005. “Predatory Trading.” *Journal of Finance*, 60(4): 1825–1863.
- Budish, Eric, Peter Cramton, and John Shim.** 2013. “The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response.” Working Paper, University of Chicago.
- Carlin, Bruce, Miguel Lobo, and S. Viswanathan.** 2007. “Episodic Liquidity Crises: Cooperative and Predatory Trading.” *Journal of Finance*, 62(5): 2235–2274.
- Chan, Louis, and Josef Lakonishok.** 1995. “The Behavior of Stock Prices Around Institutional Trades.” *Journal of Finance*, 50(4): 1147–1174.
- Duffie, Darrell.** 2010. “Presidential Address: Asset Price Dynamics with Slow-Moving Capital.” *Journal of Finance*, 65(4): 1237–1267.
- Dufour, Alfonso, and Robert Engle.** 2000. “Time and the Price Impact of a Trade.” *Journal of Finance*, 55(6): 2469–2498.
- Du, Songzi, and Haoxiang Zhu.** 2013. “Ex Post Equilibria in Double Auctions of Divisible Assets.” Working Paper, MIT Sloan School of Management.
- Grinold, Richard C., and Ronald N. Kahn.** 1995. *Active Portfolio Management: Quantitative Theory and Applications*. Probus Publ.
- Grossman, Sanford J., and Joseph E. Stiglitz.** 1980. “On the Impossibility of Informationally Efficient Markets.” *American Economic Review*, 70(3): 393–408.
- Grossman, Sanford J., and Merton H. Miller.** 1988. “Liquidity and Market Structure.” *The Journal of Finance*, 43(3): 617–633.
- Han, Jungsuk, and Albert S. Kyle.** 2013. “Speculative Equilibrium with Differences in Higher Order Beliefs.” Working Paper, Stockholm School of Economics.
- Harris, Milton, and Artur Raviv.** 1993. “Differences of Opinion Make a Horse Race.” 6(3): 473–506.
- Holthausen, Robert W., Richard W. Leftwich, and David Mayers.** 1990. “Large-Block Transactions, the Speed of Response, and Temporary and Permanent Stock-Price Effects.” *Journal of Finance*, 26(1): 71–95.
- Keim, Donald B., and Ananth Madhavan.** 1997. “Transaction Costs and Investment Style: An Inter-Exchange Analysis of Institutional Equity Trades.” *Journal of Financial Economics*, 46(3): 265–292.
- Keynes, John M.** 1936. *The General Theory of Employment, Interest and Money*. Palgrave Macmillan, United Kingdom.

- Kyle, Albert S.** 1985. "Continuous Auctions and Insider Trading." *Econometrica*, 53(6): 1315–1335.
- Kyle, Albert S.** 1989. "Informed Speculation with Imperfect Competition." *Review of Economic Studies*, 56: 317–355.
- Kyle, Albert S., and Anna A. Obizhaeva.** 2013. "Market Microstructure Invariance: Theory and Empirical Tests." Working Paper, University of Maryland.
- Kyle, Albert S., and Tao Lin.** 2001. "Continuous Speculation with Overconfident Competitors." Working Paper, Duke University.
- Kyle, Albert S., Anna A. Obizhaeva, and Yajun Wang.** 2013. "A Market Microstructure Theory of the Term Structure of Asset Returns." Working Paper, University of Maryland.
- Longstaff, Francis.** 2001. "Optimal Portfolio Choice and the Valuation of Illiquid Securities." *Review of Financial Studies*, 14: 407–431.
- Milgrom, Paul, and Nancy Stokey.** 1982. "Information, Trade and Common Knowledge." *Journal of Economic Theory*, 26(1): 17–27.
- Obizhaeva, Anna A., and Jiang Wang.** 2013. "Optimal Trading Strategy and Supply/Demand Dynamics." *Journal of Financial Markets*, 16(1): 1–32.
- Ostrovsky, Michael.** 2012. "Information Aggregation in Dynamic Markets with Strategic Traders." *Econometrica*, 80(6): 2595–2647.
- Perold, André.** 1988. "The Implementation Shortfall: Paper vs. Reality." *Journal of Portfolio Management*, 14(3): 4–9.
- Rostek, Marzena, and Marek Weretka.** 2012. "Price Inference in Small Markets." *Econometrica*, 80(2): 687–711.
- Scheinkman, José A., and Wei Xiong.** 2003. "Overconfidence and Speculative Bubbles." *Journal of Political Economy*, 111: 1183–1219.
- Staffs of the CFTC and SEC.** 2010a. *Preliminary Findings Regarding the Market Events of May 6, 2010*. Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues. May 18, 2010. Preliminary Report.
- Staffs of the CFTC and SEC.** 2010b. *Findings Regarding the Market Events of May 6, 2010*. Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues. September 30, 2010. Final Report.
- Townsend, Robert M.** 1983. "Forecasting the Forecasts of Others." *The Journal of Political Economy*, 546–588.
- Vayanos, Dimitri.** 1999. "Strategic Trading and Welfare in a Dynamic Market." *Review of Economic Studies*, 66(2): 219–254.

## Appendix

### Proof of Theorem 1.

For the second order condition to have the correct sign, we need to have  $\frac{2}{(N-1)\gamma} + \frac{A}{\tau} > 0$ , i.e.,

$$(92) \quad \frac{A}{\tau} \cdot \frac{N\tau_H^{1/2}}{(N-2)\tau_H^{1/2} - 2(N-1)\tau_L^{1/2}} > 0.$$

Therefore, the second order condition holds if and only if  $(N-2)\tau_H^{1/2} - 2(N-1)\tau_L^{1/2} > 0$ . Substituting (14) into (11), we obtain trader  $n$ 's optimal demand  $x_n^*$ . Substituting it into the market clearing condition  $\sum_{m=1}^N X_m(i_0, i_m, p) = 0$ , we obtain the equilibrium price  $P^*$ . Q.E.D.

### Proof of Lemma 1.

Applying the Kalman-Bucy filter to the filtering problem summarized by equation (24) for signals and by equations (25) and (26) for observations, we find that the filtering estimate is defined by the Itô differential equation

$$(93) \quad dG(t) = -\alpha_G \cdot G(t) \cdot dt + \sum_{n=0}^N \sigma_G^2 \cdot \bar{\Omega} \cdot \frac{\bar{\tau}_n^{1/2}}{\sigma_G \cdot \bar{\Omega}^{1/2}} \cdot \left( dI_n(t) - G(t) \cdot \frac{\bar{\tau}_n^{1/2}}{\sigma_G \cdot \bar{\Omega}^{1/2}} \cdot dt \right).$$

The mean-square filtering error of the estimate  $G(t)$ , denoted  $\sigma_G^2 \cdot \Omega$ , is defined by the Riccati differential equation

$$(94) \quad \sigma_G^2 \dot{\bar{\Omega}} = -2\alpha_G \cdot \sigma_G^2 \cdot \bar{\Omega} + \sigma_G^2 - \sigma_G^4 \bar{\Omega}^2 \sum_{n=0}^N \left( \frac{\bar{\tau}_n^{1/2}}{\sigma_G \bar{\Omega}^{1/2}} \right)^2.$$

Rearranging terms in the first equation yields equation (30). Using the steady-state assumption that  $\dot{\bar{\Omega}} = 0$  and solving the second equation for  $\bar{\Omega}$  yields equation (29). Q.E.D.

### Proof of Theorem 2.

Suppressing a subscript  $n$  for notational simplicity, the Hamilton-Jacobi-Bellman (HJB) equation corresponding to the conjectured value function  $V(M_n, S_n, D, \hat{H}_n, \hat{H}_{-n})$

in equation (48) is

$$(95) \quad 0 = \max_{c,x} \left\{ U(c) - \rho V + \frac{\partial V}{\partial M_n} (rM_n + S_n D - c - P(x) \cdot x) + \frac{\partial V}{\partial S_n} x \right\} \\ + \frac{\partial V}{\partial D} (-\alpha_D D + \sigma_G \Omega^{1/2} \sqrt{\tau_H} \hat{H}_n + \sigma_G \Omega^{1/2} (N-1) \sqrt{\tau_L} \hat{H}_{-n}) \\ + \frac{\partial V}{\partial \hat{H}_n} \left( -(\alpha_G + \tau) \hat{H}_n(t) + (\sqrt{\tau_H} + \hat{A} \sqrt{\tau_0}) (\sqrt{\tau_H} \hat{H}_n + (N-1) \sqrt{\tau_L} \hat{H}_{-n}) \right) \\ + \frac{\partial V}{\partial \hat{H}_{-n}} \left( -(\alpha_G + \tau) \hat{H}_{-n}(t) + (\sqrt{\tau_L} + \hat{A} \sqrt{\tau_0}) (\sqrt{\tau_H} \hat{H}_n + (N-1) \sqrt{\tau_L} \hat{H}_{-n}) \right) + \frac{1}{2} \frac{\partial^2 V}{\partial D^2} \sigma_D^2 \\ + \frac{1}{2} \frac{\partial^2 V}{\partial \hat{H}_n^2} (1 + \hat{A}^2) + \frac{1}{2} \frac{\partial^2 V}{\partial \hat{H}_{-n}^2} \left( \frac{1}{N-1} + \hat{A}^2 \right) + \left( \frac{\partial^2 V}{\partial D \partial \hat{H}_n} + \frac{\partial^2 V}{\partial D \partial \hat{H}_{-n}} \right) \hat{A} \sigma_D + \frac{\partial^2 V}{\partial \hat{H}_n \partial \hat{H}_{-n}} \hat{A}^2.$$

For the specific quadratic specification of the value function in equation (60), the Hamilton-Jacobi-Bellman (HJB) equation becomes

$$(96) \quad 0 = \min_{c,x} \left\{ -\frac{e^{-Ac}}{V} - \rho + \psi_M (rM_n + S_n \cdot D - c - P(x) \cdot x) \right. \\ \left. + (\psi_{SS} S_n + \psi_{SD} D + \psi_{Sn} \hat{H}_n + \psi_{Sx} \hat{H}_{-n}) x \right\} \\ + \psi_{SD} S_n (-\alpha_D D + \sigma_G \Omega^{1/2} \sqrt{\tau_H} \hat{H}_n + \sigma_G \Omega^{1/2} (N-1) \sqrt{\tau_L} \hat{H}_{-n}) \\ + (\psi_{Sn} S_n + \psi_{nn} \hat{H}_n + \psi_{nx} \hat{H}_{-n}) \\ \cdot \left( -(\alpha_G + \tau) \hat{H}_n(t) + (\sqrt{\tau_H} + \hat{A} \sqrt{\tau_0}) (\sqrt{\tau_H} \hat{H}_n + (N-1) \sqrt{\tau_L} \hat{H}_{-n}) \right) \\ + (\psi_{Sx} S_n + \psi_{xx} \hat{H}_{-n} + \psi_{nx} \hat{H}_n) \\ \cdot \left( -(\alpha_G + \tau) \hat{H}_{-n}(t) + (\sqrt{\tau_L} + \hat{A} \sqrt{\tau_0}) (\sqrt{\tau_H} \hat{H}_n + (N-1) \sqrt{\tau_L} \hat{H}_{-n}) \right) \\ + \frac{1}{2} \psi_{SD}^2 S_n^2 \sigma_D^2 + \frac{1}{2} \left( (\psi_{Sn} S_n + \psi_{nn} \hat{H}_n + \psi_{nx} \hat{H}_{-n})^2 + \psi_{nn} \right) \cdot (1 + \hat{A}^2) \\ + \frac{1}{2} \left( (\psi_{Sx} S_n + \psi_{xx} \hat{H}_{-n} + \psi_{nx} \hat{H}_n)^2 + \psi_{xx} \right) \cdot \left( \frac{1}{N-1} + \hat{A}^2 \right) \\ + ((\psi_{Sn} + \psi_{Sx}) S_n + (\psi_{nn} + \psi_{nx}) \hat{H}_n + (\psi_{xx} + \psi_{nx}) \hat{H}_{-n}) \cdot \psi_{SD} S_n \hat{A} \sigma_D \\ \left. + ((\psi_{Sn} S_n + \psi_{nn} \hat{H}_n + \psi_{nx} \hat{H}_{-n}) \cdot (\psi_{Sx} S_n + \psi_{xx} \hat{H}_{-n} + \psi_{nx} \hat{H}_n) + \psi_{nx}) \cdot \hat{A}^2. \right)$$

The solutions for optimal consumption is

$$(97) \quad c_n^*(t) = -\frac{1}{A} \cdot \log \left( \frac{\psi_M \cdot V(t)}{A} \right).$$

In the HJB equation (96), the price  $P(x)$  is linear in  $x$  based on equation (57). Plugging  $P(x)$  from equation (57) into the HJB equation (96) yields a quadratic function of  $x$  which captures the effect of trader  $n$ 's trading rate  $x_n$  on prices. Because the exponent of the conjectured value function is a quadratic function of the state variables, the optimal trading strategy is a linear function of the state variables given by

$$(98) \quad x_n^*(t) = \frac{(N-1)\gamma_P}{2\psi_M} \cdot \left[ \left( \psi_{SD} - \frac{\psi_M \gamma_D}{\gamma_P} \right) \cdot D(t) + \left( \psi_{SS} - \frac{\psi_M \gamma_S}{(N-1)\gamma_P} \right) \cdot S_n(t) \right. \\ \left. + \psi_{Sn} \cdot \hat{H}_n(t) + \left( \psi_{Sx} - \frac{\psi_M \gamma_H}{\gamma_P} \right) \cdot \hat{H}_{-n}(t) \right].$$

The derivation of this optimal trading strategy assumes that trader  $n$  observes the values of  $D(t)$ ,  $S_n(t)$ ,  $\hat{H}_n(t)$ , and  $\hat{H}_{-n}(t)$ . Although trader  $n$  does not actually observe  $\hat{H}_{-n}(t)$ , he can implement the optimal quantity  $x_n^*$  by submitting an appropriate linear demand schedule. We can think of this demand schedule as a linear function of  $P(t)$  whose intercept is a linear function of  $D(t)$ ,  $S_n(t)$ , and  $\hat{H}_n(t)$ . Trader  $n$  can infer from the market-clearing condition (56) that  $\hat{H}_{-n}$  is given by

$$(99) \quad \hat{H}_{-n}(t) = \frac{\gamma_P}{\gamma_H} \cdot \left( P(t) - D(t) \cdot \frac{\gamma_D}{\gamma_P} \right) - \frac{1}{(N-1)\gamma_H} \cdot x_n^*(t) - \frac{\gamma_S}{(N-1)\gamma_H} \cdot S_n(t).$$

Plugging equation (99) into equation (98) and solving for  $x_n^*(t)$  implements the optimal trading strategy  $x_n^*(t)$  as a linear demand schedule which depends on the price  $P(t)$  and state variables  $\hat{H}_n$ ,  $S_n(t)$ ,  $D(t)$ , which the trader directly observes. This schedule is given by

$$(100) \quad x_n^*(t) = \frac{(N-1)\gamma_P}{\psi_M} \cdot \left( 1 + \frac{\psi_{Sx}}{\psi_M} \frac{\gamma_P}{\gamma_H} \right)^{-1} \cdot \left[ \left( \psi_{SD} - \psi_{Sx} \frac{\gamma_D}{\gamma_H} \right) \cdot D(t) + \left( \psi_{SS} - \psi_{Sx} \frac{\gamma_S}{(N-1)\gamma_H} \right) \cdot S_n(t) + \psi_{Sn} \cdot \hat{H}_n(t) + \left( \psi_{Sx} \frac{\gamma_P}{\gamma_H} - \psi_M \right) \cdot P(t) \right].$$

Symmetry requires that this demand schedule be the same as the demand schedule conjectured for the  $N-1$  other traders. Equating the coefficients of  $D(t)$ ,  $\hat{H}_n(t)$ ,  $S_n(t)$ , and  $P(t)$  in equation (100) to the conjectured coefficients  $\gamma_D$ ,  $\gamma_H$ ,  $-\gamma_S$ , and  $-\gamma_P$  results in the following four restrictions that the values of the  $\gamma$ -parameters and  $\psi$ -parameters must satisfy in a symmetric equilibrium with linear trading strategies:

$$(101) \quad \frac{(N-1)\gamma_P}{\psi_M} \cdot \left( 1 + \frac{\psi_{Sx}}{\psi_M} \frac{\gamma_P}{\gamma_H} \right)^{-1} \cdot \left( \psi_{SD} - \psi_{Sx} \frac{\gamma_D}{\gamma_H} \right) = \gamma_D,$$

$$(102) \quad \frac{(N-1)\gamma_P}{\psi_M} \cdot \left( 1 + \frac{\psi_{Sx}}{\psi_M} \frac{\gamma_P}{\gamma_H} \right)^{-1} \cdot \psi_{Sn} = \gamma_H,$$

$$(103) \quad \frac{(N-1)\gamma_P}{\psi_M} \cdot \left( 1 + \frac{\psi_{Sx}}{\psi_M} \frac{\gamma_P}{\gamma_H} \right)^{-1} \cdot \left( \psi_{SS} - \psi_{Sx} \frac{\gamma_S}{(N-1)\gamma_H} \right) = -\gamma_S,$$

$$(104) \quad \frac{(N-1)\gamma_P}{\psi_M} \cdot \left( 1 + \frac{\psi_{Sx}}{\psi_M} \frac{\gamma_P}{\gamma_H} \right)^{-1} \cdot \left( \psi_{Sx} \frac{\gamma_P}{\gamma_H} - \psi_M \right) = -\gamma_P.$$

Solving this system, we obtain four equations in terms of the four unknowns  $\psi_{Sx}$ ,  $\gamma_H$ ,  $\gamma_S$ , and  $\gamma_D$ . The solution is

$$(105) \quad \psi_{Sx} = \frac{N-2}{2} \psi_{Sn}, \quad \gamma_H = \frac{N\gamma_P}{2\psi_M} \psi_{Sn}, \quad \gamma_S = -\frac{(N-1)\gamma_P}{\psi_M} \psi_{SS}, \quad \gamma_D = \frac{\gamma_P}{\psi_M} \psi_{SD}.$$

Plugging the last equation into equation (98) implies that traders will not trade on public information. It is intuitively obvious that traders cannot trade on the basis of the public information  $D(t)$  because all traders would want to trade in the same direction. Substituting equation (105) into equation (98) yields the solution for optimal strategy.

$$(106) \quad x_n^*(t) = \gamma_S \cdot \left( C_L \cdot (H_n(t) - H_{-n}(t)) - S_n(t) \right).$$

One might expect that the solution of the maximization problem will yield solutions for the nine  $\psi$ -parameters as functions of the four  $\gamma$ -parameters. One might also expect that imposing symmetry by equating the four optimal  $\gamma$ -parameters (implied by trader  $n$ 's optimal trading strategy to the four conjectured  $\gamma$ -parameters will yield solutions for the four  $\gamma$ -parameters as functions of the nine  $\psi$ -parameters. In principle, one could then expect a solution to the thirteen equations in thirteen unknowns to describe a steady-state equilibrium, if one exists.

Although this is the intuition for the solution methodology, the solution does not work in this straightforward manner. The four equations for the  $\gamma$ -parameters do not determine  $\gamma_P$  as a function of the nine  $\psi$ -parameters. Instead, the solution to the four  $\gamma$ -equations implies a restriction on the  $\psi$ -parameters which must hold in a steady state equilibrium. This restriction has an interesting economic interpretation, and we discuss it in some detail in section III.

Plugging (97) and (98) back into the Bellman equation and setting the constant term and the coefficients of  $M_n$ ,  $S_n \cdot D$ ,  $S_n^2$ ,  $S_n \cdot \hat{H}_n$ ,  $S_n \cdot \hat{H}_{-n}$ ,  $\hat{H}_n^2$ ,  $\hat{H}_{-n}^2$ , and  $\hat{H}_n \cdot \hat{H}_{-n}$  to be zero, we obtain nine equations. There are in total nine equations in nine unknowns  $\gamma_P$ ,  $\psi_0$ ,  $\psi_M$ ,  $\psi_{SD}$ ,  $\psi_{SS}$ ,  $\psi_{Sn}$ ,  $\psi_{nn}$ ,  $\psi_{xx}$ , and  $\psi_{nx}$ .

By setting the constant term, coefficient of  $M$ , and coefficient of  $SD$  to be zero, we obtain

$$(107) \quad \psi_M = -rA,$$

$$(108) \quad \psi_{SD} = -\frac{rA}{r + \alpha_D},$$

$$(109) \quad \psi_0 = 1 - \log\{r\} + \frac{1}{r} \left( -\rho + \frac{1}{2}(1 + \hat{A}^2)\psi_{nn} + \frac{1}{2} \left( \frac{1}{N-1} + \hat{A}^2 \right) \psi_{xx} + \hat{A}^2 \psi_{nx} \right).$$

In addition, by setting the coefficients of  $S_n^2$ ,  $S_n \cdot \hat{H}_n$ ,  $S_n \cdot \hat{H}_{-n}$ ,  $\hat{H}_n^2$ ,  $\hat{H}_{-n}^2$  and  $\hat{H}_n \cdot \hat{H}_{-n}$  to be zero, we obtain six polynomial equations in the six unknowns  $\gamma_P$ ,  $\psi_{SS}$ ,  $\psi_{Sn}$ ,  $\psi_{nn}$ ,  $\psi_{xx}$ , and  $\psi_{nx}$ . Defining the constants  $a_1$ ,  $a_2$ ,  $a_3$ , and  $a_4$  by

$$(110) \quad \begin{aligned} a_1 &= -\alpha_G - \tau + \sqrt{\tau_H}(\sqrt{\tau_H} + \hat{A}\sqrt{\tau_0}), \\ a_2 &= -\alpha_G - \tau + (N-1)\sqrt{\tau_L}(\sqrt{\tau_L} + \hat{A}\sqrt{\tau_0}), \\ a_3 &= (\sqrt{\tau_H} + \hat{A}\sqrt{\tau_0})(N-1)\sqrt{\tau_L}, \\ a_4 &= (\sqrt{\tau_L} + \hat{A}\sqrt{\tau_0})\sqrt{\tau_H}, \end{aligned}$$

these six equations in six unknowns can be written

(111)  $S_n^2$ :

$$0 = -\frac{1}{2}r\psi_{SS} - \frac{\gamma_P(N-1)}{rA}\psi_{SS}^2 + \frac{r^2 A^2 \sigma_D^2}{2(r+\alpha_D)^2} + \frac{1}{2}(1+\hat{A}^2)\psi_{Sn}^2 \\ + \frac{1}{2} \left( \frac{1}{N-1} + \hat{A}^2 \right) \frac{(N-2)^2}{4} \psi_{Sn}^2 - \frac{rA}{r+\alpha_D} \hat{A} \sigma_D \frac{N}{2} \psi_{Sn} + \hat{A}^2 \frac{N-2}{2} \psi_{Sn}^2,$$

(112)  $S_n \hat{H}_n$ :

$$0 = -r\psi_{Sn} - \frac{\gamma_P(N-1)}{rA}\psi_{SS}\psi_{Sn} - \frac{rA}{r+\alpha_D}\sigma_G \Omega^{1/2} \sqrt{\tau_H} + a_1 \psi_{Sn} \\ + \frac{N-2}{2} a_4 \psi_{Sn} + (1+\hat{A}^2)\psi_{nn}\psi_{Sn} + \frac{N-2}{2} \left( \frac{1}{N-1} + \hat{A}^2 \right) \psi_{nx}\psi_{Sn} \\ - \frac{rA}{r+\alpha_D} \hat{A} \sigma_D (\psi_{nn} + \psi_{nx}) + \hat{A}^2 \psi_{nx}\psi_{Sn} + \frac{N-2}{2} \hat{A}^2 \psi_{nn}\psi_{Sn},$$

(113)  $S_n \hat{H}_{-n}$ :

$$0 = -r \frac{N-2}{2} \psi_{Sn} + \frac{\gamma_P(N-1)}{rA} \psi_{SS}\psi_{Sn} - \frac{rA}{r+\alpha_D} \sigma_G \Omega^{1/2} (N-1) \sqrt{\tau_L} \\ + \left( a_3 + \frac{N-2}{2} a_2 \right) \psi_{Sn} + (1+\hat{A}^2)\psi_{Sn}\psi_{nx} + \frac{N-2}{2} \left( \frac{1}{N-1} + \hat{A}^2 \right) \psi_{xx}\psi_{Sn} \\ - \frac{rA}{r+\alpha_D} \hat{A} \sigma_D (\psi_{xx} + \psi_{nx}) + \hat{A}^2 \psi_{xx}\psi_{Sn} + \frac{N-2}{2} \hat{A}^2 \psi_{nx}\psi_{Sn},$$

(114)  $\hat{H}_n^2$ :

$$0 = -\frac{r}{2}\psi_{nn} - \frac{\gamma_P(N-1)}{4rA}\psi_{Sn}^2 + a_1 \psi_{nn} + a_4 \psi_{nx} + \frac{1}{2}(1+\hat{A}^2)\psi_{nn}^2 \\ + \frac{1}{2} \left( \frac{1}{N-1} + \hat{A}^2 \right) \psi_{nx}^2 + \hat{A}^2 \psi_{nn}\psi_{nx},$$

(115)  $\hat{H}_{-n}^2$ :

$$0 = -\frac{r}{2}\psi_{xx} - \frac{\gamma_P(N-1)}{4rA}\psi_{Sn}^2 + a_2 \psi_{xx} + a_3 \psi_{nx} + \frac{1+\hat{A}^2}{2}\psi_{nx}^2 \\ + \frac{1}{2} \left( \frac{1}{N-1} + \hat{A}^2 \right) \psi_{xx}^2 + \hat{A}^2 \psi_{xx}\psi_{nx},$$

(116)  $\hat{H}_n \hat{H}_{-n}$  :

$$0 = -r\psi_{nx} + \frac{\gamma_P(N-1)}{2rA}\psi_{Sn}^2 + a_3\psi_{nn} + a_4\psi_{xx} + (a_1 + a_2)\psi_{nx} \\ + (1 + \hat{A}^2)\psi_{nn}\psi_{nx} + \left(\frac{1}{N-1} + \hat{A}^2\right)\psi_{xx}\psi_{nx} + \hat{A}^2(\psi_{nn}\psi_{xx} + \psi_{nx}^2).$$

If this system of equations has a solution, then the solution defines a “flow equilibrium” with symmetric linear trading strategies if the second order condition and transversality condition hold. We find that solution numerically.

Note that there is always a trivial no-trade equilibrium, as in one-period model. If each trader submits a demand schedule  $X_n(t, \cdot) \equiv 0$ , then such a no-trade demand schedule is optimal for all traders. This is not a symmetric linear equilibrium in which an auctioneer can establish a meaningful market price.

The transversality condition is equivalent to  $r > 0$ : From the HJB equation and equations (111)-(116), we have

$$(117) \quad E_t^n \{dV(M_n(t), S_n(t), D(t), \hat{H}_n(t), \hat{H}_{-n}(t))\} = \\ -(r - \rho) \cdot V(M_n(t), S_n(t), D(t), \hat{H}_n(t), \hat{H}_{-n}(t)) \cdot dt.$$

This yields

$$(118) \quad E_t^n \{e^{-\rho(T-t)}V(M_n(T), S_n(T), D(T), \hat{H}_n(T), \hat{H}_{-n}(T))\} = \\ e^{-r(T-t)} \cdot V(M_n(t), S_n(t), D(t), \hat{H}_n(t), \hat{H}_{-n}(t)),$$

which implies that the transversality condition (54) is satisfied if  $r > 0$ .

The second order condition requires  $\gamma_P > 0$ . For the minimum in the optimization problem (96) to exist, the second order condition requires the  $2 \times 2$  matrix

$$(119) \quad \begin{pmatrix} -\frac{A^2}{V} & 0 \\ 0 & \frac{2 \cdot r \cdot A}{(N-1)\gamma_P} \end{pmatrix}$$

be positive definite. Since value function  $V$  is negative, this condition holds when demand schedules are downward sloping with  $\gamma_P > 0$ .

In addition to  $\gamma_P > 0$ , intuition suggests that the verification principle requires equilibrium permanent price impact to have the correct sign  $\gamma_S > 0$ . If  $\gamma_S < 0$ , this would imply  $\psi_{SS} < 0$  (wrong sign for value of inventories) from equation (61). All traders would believe they could achieve infinite utility by purchasing (selling) large quantities at lower and lower (higher and higher) prices. This is incompatible with an equilibrium with linear strategies of the conjectured form. Similarly, equilibrium requires  $\psi_{Sn} < 0$ , implying  $\gamma_H > 0$ . Q.E.D.