

Centre for
Economic
and Financial
Research
at
New Economic
School



February 2016

Market Microstructure Invariance: A Dynamic Equilibrium Model

Albert S. Kyle
Anna A. Obizhaeva

Working Paper No 228

CEFIR / NES Working Paper series

Market Microstructure Invariance: A Dynamic Equilibrium Model

Albert S. Kyle
University of Maryland
akyle@rhsmith.umd.edu

Anna A. Obizhaeva
New Economic School
aobizhaeva@nes.ru

First Draft: January 17, 2016
This Draft: February 8, 2016*

Abstract

We derive invariance relationships for a dynamic infinite-horizon model of market microstructure with risk-neutral informed trading, noise trading, market making, and endogenous production of information. Equilibrium prices follow a martingale with endogenously derived stochastic volatility. The invariance relationships for bet sizes and transaction costs are obtained under the assumption that the effort required to generate one discrete bet does not vary across securities and time. The invariance relationships for pricing accuracy and market resiliency require the additional assumption that private information has the same signal-to-noise ratio across markets. Since bets are based on the arrival of discrete chunks of information, the structural model describes how the invariance relationships reflect differences in the granularity of information flows across markets. The model links proportionality coefficients in invariance relationships to fundamental parameters.

Keywords: market microstructure, invariance, liquidity, bid-ask spread, market impact, transaction costs, market efficiency, efficient markets hypothesis, pricing accuracy, resiliency, order size.

*This paper was previously part of the manuscript “Market Microstructure Invariants: Theory and Empirical Tests” (October 17, 2014).

Kyle and Obizhaeva (2016) present market microstructure invariance as two empirical hypotheses which explain a substantial fraction of the cross-sectional variation in bet size (or meta-order size) and transaction costs across stocks. Here we develop theoretical foundations for microstructure invariance by describing a theoretical model of trading from which invariance relationships are derived endogenously. The structure of the model has the flavor of Treynor (1995) and Glosten and Milgrom (1985), with underlying principles similar to the continuous-time model of Kyle (1985).

The model is based on the following assumptions. The unobserved fundamental value of the stock follows geometric Brownian motion. Informed and noise traders arrive randomly and trade only once. Each informed trader pays a fixed dollar cost to obtain an informative private signal about the fundamental value. Each noise trader obtains a “fake” signal which has the same unconditional distribution as an informative signal but contains no information. Each noise trader believes himself to be an informed trader. Both informed traders and noise traders trade optimally on the perceived information content of their signals. Noise traders place bets at a rate which leads to an exogenously given turnover rate of the float of shares. The number of informed traders adjusts endogenously so that informed traders make zero expected profits net of the cost of a signal and the price impact of trading on it. Each informed or noise trader places one bet by announcing to competitive risk-neutral market makers the quantity he wishes to trade. Conditional on this announced quantity but not knowing whether the trader is informed or not, market makers set a break-even price and hold the resulting inventory until the game is over.

Our main contribution is to derive market microstructure invariance hypotheses as endogenously implied properties of this well-specified theoretical model. According to the invariance hypotheses, microstructure characteristics related to the size, frequency, and market impact cost of bets become constants when scaled in units of business time measured as the rate at which bets arrive into the market. Under the key assumption that the expected dollar market impact cost of executing a bet is constant across markets or across time, we show that the following two empirical hypotheses conjectured by Kyle and Obizhaeva (2016) indeed hold in the equilibrium of the theoretical model. First, the hypothesis of *bet size invariance* says that the distribution of the dollar risk transferred by all bets is the same when the dollar risk is measured in units of business time. Second, the hypothesis of *transaction costs invariance* says that the expected dollar transaction cost of executing a bet is the same function of the size of the bet when the bet’s size is measured as the dollar risk it transfers in units of business time. Furthermore, under the additional assumption of constant informativeness of the signals upon which bets are based, we also derive additional invariance principles which we call *pricing accuracy invariance* and *market resiliency invariance*. Pricing accuracy is invariant across markets if its reciprocal (pricing error) is scaled by returns volatility per unit of business time, and market resiliency is invariant if it is measured in units of business time.

The main theorem (see equations (37)) summarizes the scaling laws for bet size,

number of bets, market depth, market liquidity, pricing accuracy, and market resiliency. These scaling laws connect market microstructure variables to trading volume and volatility using specific exponents of one-third and two-thirds.

The seemingly obscure invariance relationships are in fact the result of several properties shared by many models of market microstructure (see equations (38)–(41)). First, trading volume is defined as the sum of bets. Second, prices change only as a result of the information content of the order flow, consistent with the Wall Street adage “it takes volume to make prices move.” The model has linear price impact which links return volatility to the information content of bets. Third, the dollar cost of acquiring an informative signal, which in equilibrium equals the dollar price impact cost of the bet, is the same for all assets and across time. This key economic assumption is motivated by the intuition that asset managers allocate scarce intellectual resources across markets and across time to maximize the profitability of trading on private information. Fourth, there is a technical equation which relates two moments of the bet size distribution. The general nature of these assumptions suggests that invariance relationships are most likely not only features of a particular equilibrium model in this paper but also universal properties of financial markets.

The intuition behind the exponents of one-third and two-thirds comes from a simple observation. Both volume and returns variance are linear in the number of bets, but volatility is linear in *square root* of the number of bets. Consider what happens when trading volume increases, either as a result of a gradual time series increase in market capitalization (holding turnover approximately constant) or as a result of a cross-sectional comparison of a high-volume stock with a low-volume stock. Suppose that returns volatility remains constant and is equal to fundamental volatility. When volume is higher, both the number of bets and the size of bets are larger. When the number of bets increases, the market becomes more efficient; the distance between prices and fundamentals shrinks, market depth increases, and traders must execute larger bets to cover the exogenous fixed costs of generating each signal. Bet sizes are inversely proportional to returns volatility per bet; therefore, as markets become deeper, faster, and more efficient, bet size increases half as fast as the rate as number of bets. This leads to one-to-two ratio between the size of bets and the number of bets, implying scaling exponents of one-third and two-thirds.

The dynamic equilibrium of the model is interesting in itself. The equilibrium price process is approximately a diffusion when bet size is small. Although fundamental volatility is constant, return volatility is actually stochastic. When return volatility is less than the volatility of fundamentals, trading is not incorporating information into prices as fast as fundamental uncertainty is unfolding, the pricing error is widening, the market is becoming less liquid and less efficient, and bets are becoming smaller and more frequent. When return volatility is greater than the volatility of fundamentals, trading is incorporating information into prices faster than fundamental uncertainty is unfolding, the pricing error is shrinking, the market is becoming more liquid and more efficient, and bets are becoming larger and less frequent. A conditional steady state

is reached when returns volatility coincides with the volatility of fundamentals, and thus the rate at which new fundamental volatility unfolds and the rate at which prices incorporate private information are in balance. The endogenous returns volatility increases in the size of the pricing error and the number of traders.

In the financial economics literature, there is a long tradition of empirical studies which relate trading volume and the number of trades to return volatility. For example, Clark (1973), Tauchen and Pitts (1983), Karpoff (1987), and Harris (1987) all find that volatility is closely related to trading. Andersen et al. (2015) investigate a different volume-volatility relationship based on the hypothesis that trades in the S&P E-mini futures contract are proportional to bets. The invariance hypothesis can be reframed to predict that returns volatility per trade is inversely proportional to average trade size; this prediction is supported by the futures trading data.

The model explicitly highlights difference between existing definitions of market efficiency. The model assumes that the market is efficient in a sense that prices follow a martingale, as discussed in Fama (1970) and LeRoy (1989). Black (1986) proposes an alternative measure of market efficiency based on the size of pricing errors, defined as the log-standard deviation of the ratio of price to fundamental value. The model implies that the sizes of pricing errors are proportional to market resiliency, defined as the speed with which prices recover after a random shock. Since market resiliency may be easier to identify empirically than hard-to-observe pricing errors, this link provides an indirect way to implement this alternative measure of market efficiency.

The rest of this paper presents the model, solves for the equilibrium price process, derives invariance relationships, discusses the model's intuition, and makes several general observations about properties of the model.

1 A Dynamic Model of Trading

We next describe a dynamic model of sequential speculative trading. To make clear the distinction between endogenous variables and exogenous parameters, we use notation which assumes that all exogenous parameters are constants, while all endogenous variables are time-varying.

Fundamental Value and Prices. Let the unobserved fundamental value of the asset follow a geometric Brownian motion given by

$$F(t) := \exp[\sigma_F \cdot B(t) - \frac{1}{2} \cdot \sigma_F^2 \cdot t], \quad (1)$$

where $B(t)$ follows standardized Brownian motion with $\text{var}\{B(t + \Delta t) - B(t)\} = \Delta t$, the constant σ_F measures the “Black-Scholes volatility” of the fundamental value, and the term $\frac{1}{2} \cdot \sigma_F^2 \cdot t$ makes $F(t)$ follow an exponential martingale. The fundamental value $F(t)$ is the value to which a stock price would converge if traders continuously invested huge resources to acquire information about its value.

Trading takes place over some long finite horizon, ending at a distant date at which all traders receive a payoff equal to the fundamental value of the asset. Assume a zero risk-free rate of interest.

The price changes as informed traders and noise traders arrive into the market and anonymously place buy and sell bets. Risk-neutral competitive market makers clear the market by taking the other side of bets. They set the market price $P(t)$ as the conditional expectation of the fundamental value $F(t)$ given a history of the bet flow, which turns out to be informationally equivalent to the history of prices.

Rather than focusing on the market makers' estimate of the fundamental value $F(t)$, it is simpler to focus instead on their estimate of the Brownian motion $B(t)$, in terms of which $F(t)$ is defined in equation (1). Let $\bar{B}(t) := E_t\{B(t)\}$ denote the market's conditional expectation of $B(t)$ based on observing the history of bets or prices up to time t . For now, assume that the error $B(t) - \bar{B}(t)$ has approximately a normal distribution with a mean of zero and a time-varying variance denoted $\Sigma(t)/\sigma_F^2$. Scaling the error variance by σ_F^2 gives $\Sigma(t)^{1/2}$ the interpretation as a percentage error, consistent with the definition of pricing accuracy.

The price $P(t)$ is the best estimate of fundamental value when it is given by

$$P(t) = E_t\{F(t)\} = \exp[\sigma_F \cdot \bar{B}(t) + \frac{1}{2} \cdot \Sigma(t) - \frac{1}{2} \cdot \sigma_F^2 \cdot t]. \quad (2)$$

Given $\Sigma(t)$, this equation shows that $P(t)$ and $\bar{B}(t)$ are informationally equivalent.¹ It can be easily shown that the price has a martingale property.

Pricing Accuracy, Resiliency, and Market Efficiency. The definition of $\Sigma(t)$ implies

$$\Sigma(t) = \text{var}_t\{\ln[F(t)/P(t)]\}. \quad (3)$$

Thus, $\Sigma^{1/2}$, which measures the standard deviation of this log-difference, can be interpreted intuitively as a log percentage-error, and its reciprocal $\Sigma^{-1/2}$ can be interpreted as a measure of *pricing accuracy*.

Intuitively, the unobservable fundamental value $F(t)$ evolves over time due to continuous un-modeled changes in production processes, consumer tastes, costs of materials, prices of outputs, competitor strategies, and other market conditions. The observable price $P(t)$ is the market's best estimate of this fundamental value. If prices are above fundamentals, then prices will tend to decrease over time towards fundamentals, as trading based on private information gradually incorporates information about fundamental value into prices. If prices are below fundamentals, then informed trading will tend to make prices increase over time in the direction of fundamentals. On average, the more prices deviate from fundamental value, the less accurate they

¹Note that if $\Sigma(t) = 0$, then equations (1) and (2) are the same; the term $\frac{1}{2} \cdot \Sigma(t)$ is the convexity adjustment $\exp()$ function implied by a normal distribution. The term $\frac{1}{2} \cdot \sigma_F^2 \cdot t$ is an analogous adjustment in both equation (1) and (2). This same adjustment shows up in the moment generating function for a normal distribution.

are. The variable $\Sigma^{1/2}(t)$ measures the average log-distance between fundamental values and prices, i.e. pricing error equal to the inverse of pricing accuracy. The pricing error increases over time if innovations in the Brownian motion $B(t)$ add variance faster than it is reduced by the information inferred from informed trading.

Black (1986) uses a similar concept of pricing accuracy when he defines market efficiency. He conjectures that “almost all markets are efficient” in the sense that “price is within a factor 2 of value” at least 90% of the time. Since the probability that a normal distribution is within 1.64 standard deviations of its mean is approximately equal to 90%, Black’s conjecture holds using our notation if $\Sigma^{1/2}(t) < \ln(2)/1.64$ in almost all markets. Black would say that the market has become more “efficient,” if $\Sigma^{1/2}(t)$ becomes smaller, implying the average log-distance between observable prices and unobservable fundamentals shrinks.

It is convenient to think about pricing accuracy in time units. Let the pricing error variance $\Sigma(t)$ be scaled by annual returns variance $\sigma^2(t)$. The quantity $\Sigma(t)/\sigma^2(t)$ can be then interpreted as the number of years by which the informational content of prices lags behind fundamental value based on current returns volatility. The less accurate prices are ($\Sigma(t)$ is large) and the lower is returns volatility (σ^2 is small), the more years it takes for market prices to catch up with fundamentals. For example, suppose a stock’s annual volatility is about 35%, implying $\sigma = 0.35$, and suppose $\Sigma^{1/2} = \ln(2)/1.64$, as before. Then prices are about 1.50 years *behind* fundamental value, since $(\ln(2)/1.64)^2/0.35^2 \approx 1.50$. This means that, on average, it would take about 1.50 years of 35% returns volatility for prices to converge to fundamental value under the assumption that the fundamental value remained frozen in time. Of course, fundamental value continues to change with variance σ_F^2 , implying $\Sigma(t) \approx 1.50 \cdot \sigma_F^2$ after about 1.50 years. If $\sigma_F < 0.35$, then pricing accuracy would have improved after about 1.50 years.

The concept of *market resiliency* is closely related to the concept of pricing accuracy. They are two sides of the same coin. Let resiliency ρ be the mean-reversion parameter (per calendar year) measuring the speed with which a random shock to prices—resulting from execution of an uninformative bet—dies out over time as informative bets drive prices back towards fundamental value. The half-life of an uninformative shock to prices is then equal to $\rho^{-1} \cdot \ln(2)$. We will later show that market resiliency is greater in markets with higher pricing accuracy.

Black’s definition of market efficiency, based on accuracy of prices, contrasts sharply with the conventional definition associated with Eugene Fama. Fama considers a market to be efficient if all available information is appropriately reflected in price; this implies that prices adjusted for dividends and risk premium follow a martingale, regardless of how much information is available overall in the market. In an efficient market, the log-distance between prices and fundamentals $\Sigma^{1/2}$ may be either large or small. In our model, prices are by assumption efficient in the sense of Fama, but pricing accuracy and resiliency vary endogenously over time.

Informed Traders. There are informed traders, noise traders, and market makers. Informed traders arrive randomly into the market at endogenously determined rate $\gamma_I(t)$. Each informed trader pays a fixed cash amount for a private signal, constructs his estimate of the fundamental value, and trades on his estimate optimally maximizing profits taking into account how his trading affects prices.

If an informed trader arrives at time t , he observes one private signal $\tilde{i}(t)$ and also the history of prices, including the most recent price; he can thus infer the market's estimate $\bar{B}(t)$ and the scaled error variance $\Sigma(t)$. An informed trader then places one and only one bet. The bet quantity is announced to market makers, and the market makers take the other side of the bet at a competitive price described below. Informed signals are assumed to have the signal-plus-noise form

$$\tilde{i}(t) := \tau^{1/2} \cdot \Sigma(t)^{-1/2} \cdot \sigma_F \cdot [B(t) - \bar{B}(t)] + \tilde{Z}_I(t), \quad (4)$$

where τ is an exogenous constant parameter measuring the precision of the signal and the noise term $\tilde{Z}_I(t) \sim N(0, 1)$ is distributed independently from the history of prices prior to the arrival of the signal. The signal $\tilde{i}(t)$ is the weighted sum of two random variables, $\Sigma(t)^{-1/2} \cdot \sigma_F \cdot [B(t) - \bar{B}(t)]$ and $\tilde{Z}_I(t)$, each of which is (approximately) distributed $N(0, 1)$. The weighting factor τ , which governs the weight on the information-carrying component, measures the precision of the signal in units “per time.” Thus, scaling $\tau^{1/2}$ by the factor $\Sigma(t)^{-1/2} \cdot \sigma_F$ —the reciprocal of the standard deviation of $B(t) - \bar{B}(t)$ —makes the parameter τ measure the signal-to-noise ratio, which is assumed to be constant regardless of the level of $B(t) - \bar{B}(t)$ and price dynamics. Without scaling, signals would be much more informative when $B(t) - \bar{B}(t)$ is large and much less informative when $B(t) - \bar{B}(t)$ is small.

Since we consider a continuous-time model, we assume τ is small enough so that $\text{var}\{\tilde{i}(t)\} = 1 + \tau \approx 1$. Although the signal associated with the n th bet to arrive is different from previous signals, we omit a clarifying subscript n on $\tilde{i}(t)$.

An informed trader infers his prior estimate $\bar{B}(t)$ of $B(t)$ from market prices. Upon observing a signal $\tilde{i}(t)$, he then updates his prior from $\bar{B}(t)$ to $\bar{B}(t) + \Delta\bar{B}_I(t)$. Here the subscript I reminds that updating is performed given the information set of an informed trader. The updating by the informed trader is different from the updating by market makers (described below) because the informed trader assumes that he himself is informed with probability one while market makers—not knowing whether a trade is informed or uninformed—must incorporate into their updating the possibility that a trade comes from a noise trader. Assuming τ is small, the informed trader's update is given by

$$\Delta\bar{B}_I(t) := E\{B(t) - \bar{B}(t)|\tilde{i}(t)\} \approx \tau^{1/2} \cdot \Sigma(t)^{1/2} \cdot \sigma_F^{-1} \cdot \tilde{i}(t). \quad (5)$$

This is the valuation update rule for each informed trader, who uses his private signal $\tilde{i}(t)$ to adjust his prior estimate $\bar{B}(t)$. To prove this equation, we regress $B(t) - \bar{B}(t)$ on signal $\tilde{i}(t)$ from equation (4), plug the variance of $B(t) - \bar{B}(t)$ equal to $\Sigma(t)/\sigma_F^2$, and use the continuous-time approximation $\text{var}\{\tilde{i}(t)\} \approx 1$.

For what follows, it is useful to calculate how dollar price would change, if the signal value $\Delta\bar{B}_I(t)$ were to be fully incorporated into prices. We obtain

$$\begin{aligned} E\{F(t) - P(t) \mid \Delta\bar{B}_I(t)\} &= P(t) \cdot E\{\exp[\sigma_F \cdot (B(t) - \bar{B}(t)) - \frac{1}{2} \cdot \Sigma(t)] - 1 \mid \Delta\bar{B}_I(t)\} \\ &\approx P(t) \cdot (\exp[\sigma_F \cdot (\Delta\bar{B}_I(t) - \frac{1}{2} \cdot \Delta\bar{B}_I(t)^2)] - 1) \\ &\approx P(t) \cdot \sigma_F \cdot \Delta\bar{B}_I(t). \end{aligned} \quad (6)$$

Here we use equations (1) and (2) for $F(t)$ and $P(t)$ in the first line, the continuous-time approximation for conditional variance

$$\text{var}\{\sigma_F \cdot (B(t) - \bar{B}(t)) \mid \Delta\bar{B}_I(t)\} \approx \Sigma(t) - \Delta\bar{B}_I(t)^2 \quad (7)$$

in the second line, and a Taylor approximation in the third line.

Upon obtaining a private signal, an informed trader arriving at time t has to decide on how to trade on it. We assume that an informed trader executes a bet of size $\tilde{Q}(t)$. In order to break even in the presence of adverse selection, market makers take the other side of the bet at a price adjusted by $\lambda(t) \cdot \tilde{Q}(t)$, where $\lambda(t)$ is a price impact parameter. The price impact is linear in the size of the bet. Then, the expected dollar costs of executing $\tilde{Q}(t)$ shares is equal to

$$C_B(\tilde{Q}(t)) = \lambda(t) \cdot \tilde{Q}(t)^2. \quad (8)$$

Suppose that a bet of size $\tilde{Q}(t)$ is a linear multiple of $\Delta\bar{B}_I(t)$ in such a way that only a fraction θ of the “fully revealing” impact $P(t) \cdot \sigma_F \cdot \Delta\bar{B}_I(t)$ from equation (6) is incorporated into prices, i.e.,

$$\tilde{Q}(t) = \theta \cdot \lambda(t)^{-1} \cdot P(t) \cdot \sigma_F \cdot \Delta\bar{B}_I(t). \quad (9)$$

If the informed trader were to trade $\tilde{Q}(t)$ shares defined in equation (9) at price $P(t)$ with no price impact costs, then his unconditional expected “paper trading” profits, denoted $\bar{\pi}_I(t)$, would be equal to

$$\bar{\pi}_I(t) := E\{[F(t) - P(t)] \cdot \tilde{Q}(t)\} = \frac{\theta \cdot P(t)^2 \cdot \sigma_F^2 \cdot E\{\Delta\bar{B}_I(t)^2\}}{\lambda(t)}. \quad (10)$$

The expected profits of the informed trader $[F(t) - P(t)] \cdot \tilde{Q}(t)$ net of transactions costs $\lambda(t) \cdot \tilde{Q}(t)^2$ conditioning on observing the signal $\tilde{i}(t)$, or equivalently on observing $\Delta\bar{B}_I(t)$, are equal to

$$E\{[F(t) - P(t)] \cdot \tilde{Q}(t) - \lambda(t) \cdot \tilde{Q}(t)^2 \mid \Delta\bar{B}_I(t)\} = \frac{\theta \cdot (1 - \theta) \cdot P(t)^2 \cdot \sigma_F^2 \cdot \Delta\bar{B}_I(t)^2}{\lambda(t)}. \quad (11)$$

It is clear that

$$\theta = 1/2 \quad (12)$$

maximizes the expected profits in equation (11) and therefore solves the optimization problem of a risk-neutral monopolistic informed trader, who optimally incorporates half of his information into prices.

In what follows, rather than assuming that informed traders are risk neutral and therefore $\theta = 1/2$ as the equilibrium choice of θ , we will instead allow θ to have an arbitrary exogenous value such that $0 < \theta < 1$. This approach accommodates the possibility that informed traders are risk averse, in which case $\theta < 1/2$ might be optimal; it also accommodates the possibility of information leakage, in which case $\theta > 1/2$ might be optimal. More importantly, it allows us to show that the invariance results derived below depend only on the fact that θ is some constant, not that θ has the specific value $1/2$.

Noise Traders. Noise traders arrive randomly and place bets in the market at an endogenously determined rate $\gamma_U(t)$ such that a constant fraction η of the market capitalization of the firm turns over, on average, per day. In the spirit of Black (1986), each noise trader trades based on a “fake” signal which has the same unconditional distribution as an informative signal but no information content.

More formally, if a noise trader arrives at time t , he receives a signal $\tilde{i}(t)$ which is assumed to have the same unconditional distribution as an informative signal (4), i.e.,

$$\tilde{i}(t) = \tilde{Z}_U(t), \quad (13)$$

where $\tilde{Z}_U(t) \sim N(0, 1 + \tau) \approx N(0, 1)$. This signal is “noise” in the sense that $\tilde{i}(t)$ is distributed independently from the error $\bar{B}(t) - B(t)$ and the history of prices.

Let N denote the number of shares outstanding, let $V(t)$ denote average share volume per day from informed and noise traders combined, and let $\gamma(t) := \gamma_I(t) + \gamma_U(t)$ denote the combined arrival rate of bets by informed traders and noise traders. Both informed trades and noise trades are distributed as the random variable $\tilde{Q}(t)$ in equation (9). Then, expected share volume from noise traders $\eta \cdot N$ and the total volume $V(t)$ satisfy

$$\gamma_U(t) \cdot E\{|\tilde{Q}(t)|\} = \eta \cdot N, \quad \gamma(t) \cdot E\{|\tilde{Q}(t)|\} = V(t). \quad (14)$$

As in market microstructure invariance, the expected arrival rate of bets $\gamma(t)$ sets the pace of business time in the model.

Market Makers (Intermediaries). Zero-profit, risk-neutral, competitive market makers set prices such that the price impact of anonymous trades by either informed or noise traders make the price $P(t)$ equal to the conditional expectation of the fundamental value $F(t)$ given the history of all bets. Market makers use the linear price-adjustment rule,

$$\Delta P(t) := E\{F(t) - P(t) | \tilde{Q}(t)\} \approx \lambda(t) \cdot \tilde{Q}(t). \quad (15)$$

If market makers could observe whether a bet was placed by an informed trader or noise trader, they would multiply the price impact of informed bets $\lambda(t) \cdot \tilde{Q}(t)$ by $1/\theta$ (since informed bets reveal only fraction θ of their information content by assumption) and they would multiply the price impact of noise bets $\lambda(t) \cdot \tilde{Q}(t)$ by zero (since noise bets have no information content). The probability of an informed bet is $\gamma_I(t)/\gamma(t)$, and the probability of a noise bet is $\gamma_U(t)/\gamma(t)$. In equilibrium, the average impact of a bet must satisfy

$$\Delta P(t) = \lambda(t) \cdot \tilde{Q}(t) = \frac{\gamma_I(t)}{\gamma(t)} \cdot \lambda(t) \cdot \tilde{Q}(t) \cdot \frac{1}{\theta} + \frac{\gamma_U(t)}{\gamma(t)} \cdot \lambda(t) \cdot \tilde{Q}(t) \cdot 0. \quad (16)$$

This equation has the interesting property that price impact $\lambda(t)$ and $\tilde{Q}(t)$ enters both of its sides linearly. This implies that the market makers' updating rule does not define the market depth $1/\lambda(t)$ but instead imposes a restriction on the fraction of informed traders and noise traders given by

$$\frac{\gamma_I(t)}{\gamma(t)} = 1 - \frac{\gamma_U(t)}{\gamma(t)} = \theta. \quad (17)$$

The endogenously determined fraction of traders who are informed turns out to be equal to the exogenous constant θ . Recall that in the special case when informed traders are risk neutral, $\theta = 1/2$ implies that the number of informed traders is exactly equal to the number of noise traders. Thus, the fraction of informed trades is not determined by equation (11) describing the profit optimization of informed traders described but rather by equation (16) describing the adverse selection problem faced by market makers. As discussed in section below, this property is not unique for our model; for example, it can be also found in the derivation of the continuous-time model of Kyle (1985).

Equations (14) and (17) further imply that, in terms of exogenous variables, share volume $V(t)$ constitutes a constant fraction of shares outstanding N ,

$$V(t) = \eta \cdot N / (1 - \theta). \quad (18)$$

Break-Even Conditions for Market Makers and Informed Traders. We will first derive the break-even condition for market makers. Equation (9) implies that the unconditional expected price impact cost of an informed bet, denoted $\bar{C}_B(t)$, is given by

$$\bar{C}_B(t) := E\{C_B(\tilde{Q}(t))\} = \lambda(t) \cdot E\{\tilde{Q}(t)^2\} = \frac{\theta^2 \cdot P(t)^2 \cdot \sigma_F^2 \cdot E\{\Delta \bar{B}_I(t)^2\}}{\lambda(t)}. \quad (19)$$

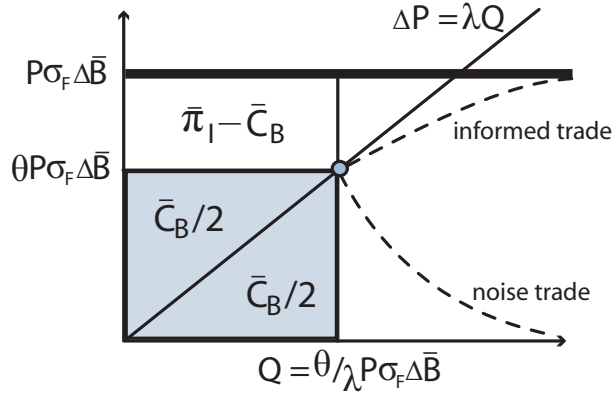
Bets by noise traders have the same expected impact cost since they have the same unconditional distribution as informed bets and are indistinguishable from informed bets from the perspective of market makers. The equilibrium level of costs must allow

market makers to break even. Thus, the expected dollar price impact costs $\bar{C}_B(t)$ that market makers expect to collect from all $\gamma_I(t) + \gamma_U(t)$ traders must be equal to the expected dollar paper trading profits $\bar{\pi}_I(t)$ of γ_I informed traders:

$$(\gamma_I(t) + \gamma_U(t)) \cdot \bar{C}_B(t) = \gamma_I \cdot \bar{\pi}_I(t). \quad (20)$$

Figure 1 illustrates the intuition informally. Informed traders incorporate only a fraction θ of their information into prices by trading $\tilde{Q}(t)$ and the price jumps by $\theta \cdot E\{F(t) - P(t) | \tilde{Q}(t)\}$; informed traders pay transactions costs $\bar{C}_B(t)$ and expect to make $\bar{\pi}_I(t) - \bar{C}_B(t)$ as the price gradually converges to expected fundamental value $E\{F(t) | \tilde{Q}(t)\}$ over time due to the subsequent trading of other informed traders. These profits are realized at some distant date when the game ends and positions are liquidated at the expected fundamental value $F(t)$. Noise traders execute orders which incur transactions costs $\bar{C}_B(t)$ but would earn nothing if there were no transactions costs. As in Treynor (1995), the expected losses market makers incur trading with informed traders $\gamma_I(t) \cdot (\bar{\pi}_I(t) - \bar{C}_B(t))$ are equal to their expected gains trading with noise traders $\gamma_U(t) \cdot \bar{C}_B(t)$. This is an alternative explanation for equation (20).

Figure 1: Intuition for the Model.



There is price continuation after an informed trade and mean reversion after a noise trade.

We next derive the break-even condition for informed traders. The break-even condition for informed traders yields the rate at which informed traders place bets $\gamma_I(t)$. The expected paper trading profits from trading on a signal $\bar{\pi}_I(t)$ must equal the sum of expected transaction costs $\bar{C}_B(t)$ and the exogenously constant cost of acquiring private information denoted c_I at the first place,

$$\bar{\pi}_I(t) = \bar{C}_B(t) + c_I. \quad (21)$$

Liquidity Measure. Our main measure of liquidity is the expected percentage cost of executing a bet (in basis points), defined as

$$1/L(t) := \frac{\bar{C}_B(t)}{E\{|P(t) \cdot \tilde{Q}(t)|\}}. \quad (22)$$

It measures the dollar-volume-weighted expected percentage cost of executing a bet and intuitively expresses the expected dollar transaction cost as a fraction of the expected dollar value traded.

Diffusion Approximation. The model theoretically implies that the price follows a jump process which changes when a bet arrives. Conditional expectations of $\bar{B}(t)$ calculated based on a random number of bet arrivals during a given time interval are not precisely linear in their size. Moreover, since the price is a nonlinear function of $\bar{B}(t)$, price impact is theoretically nonlinear as well.

To deal with these non-linearities, we assume that the arrival rate of bets is so fast—and the resulting price impact of each bet is so small—that a linear approximation based on a diffusion is appropriate. In the limit as the bet arrival rate $\gamma(t)$ goes to infinity, the conditional expectation $\bar{B}(t)$ becomes an exactly linear function of the history of bets, and the price process becomes a diffusion. This is compatible with assuming that market makers and traders use linear filtering, market makers offer linear supply schedules to traders, traders place bets as linear functions of signals, log-price changes are normally distributed, and price changes are conditionally normally distributed in the sense of an Euler approximation.

As a result of each bet, market makers update their estimate of $\bar{B}(t)$ by $\Delta\bar{B}(t)$. This update of market makers is not to be confused with the update $\Delta\bar{B}_I(t)$ of informed traders, since they use different information set.

A trade is informed with probability θ ; if informed, a trade incorporates a fraction θ of its information content into prices and leads to an adjustment in market maker's estimate $\bar{B}(t)$ of

$$\Delta\bar{B}(t) = \theta \cdot \tau^{1/2} \cdot \Sigma(t)^{1/2} \cdot \sigma_F^{-1} \cdot \left(\tau^{1/2} \cdot \Sigma(t)^{-1/2} \cdot \sigma_F \cdot [B(t) - \bar{B}(t)] + \tilde{Z}_I(t) \right), \quad (23)$$

obtained from the definition of informed traders' signal $\tilde{i}(t)$ in equations (4) and the filtering rule in equation (5). A trade is uninformed with probability $1 - \theta$; if uninformed, a trade adds to market maker's estimate $\bar{B}(t)$ noise of size

$$\Delta\bar{B}(t) = \theta \cdot \tau^{1/2} \cdot \Sigma(t)^{1/2} \cdot \sigma_F^{-1} \cdot \tilde{Z}_U(t), \quad (24)$$

obtained from the definition of noise traders' signal $\tilde{i}(t)$ in equations (13) and the filtering rule in equation (5).

When the arrival rate of bets $\gamma(t)$ per day is sufficiently large, the diffusion approximation for the dynamics of the estimate $\bar{B}(t)$ is described by a mixture of $\theta \cdot \gamma(t)$

adjustments of the form (23) and $(1-\theta)\cdot\gamma(t)$ adjustments of the form (24). Combining both of them, we write the dynamics of the estimate $\bar{B}(t)$,

$$d\bar{B}(t) = \gamma(t) \cdot \theta^2 \cdot \tau \cdot [B(t) - \bar{B}(t)] \cdot dt + \gamma(t)^{1/2} \cdot \theta \cdot \tau^{1/2} \cdot \Sigma(t)^{1/2} \cdot \sigma_F^{-1} \cdot dZ(t). \quad (25)$$

The first term corresponds to the information contained in informed signals which arrive at a rate $\theta \cdot \gamma(t)$. The second term corresponds to the noise contained in all—
informed and noise—bets arriving at a rate $\gamma(t)$. Here, $\gamma(t)^{1/2} \cdot dZ(t)$ is obtained by converting the mixture of $dZ_I(t)$ and $dZ_U(t)$ —with probabilities θ and $1 - \theta$, respectively—into a Brownian motion $dZ(t)$ with the same variance.

Conditional Steady State. The steady state of this model has interesting properties. In the equilibrium, as we show next, the log-price process is a martingale with stochastic volatility.

Define $\sigma(t)$ by

$$\sigma(t) := \theta \cdot \tau^{1/2} \cdot \Sigma(t)^{1/2} \cdot \gamma(t)^{1/2}. \quad (26)$$

By applying Ito's lemma to the definition of price in equation (2) to get $dP(t)$, plugging equation (25) into $d\bar{B}(t)^2$, and using equation (26), we obtain

$$\frac{dP(t)}{P(t)} = \frac{1}{2} \cdot [\Sigma'(t) - \sigma_F^2 + \sigma(t)^2] \cdot dt + \sigma_F \cdot d\bar{B}(t). \quad (27)$$

The competition between market markets leads to market efficiency which implies that $P(t)$ must follow a martingale. The drift in equation (27) therefore must be zero, implying that $\Sigma(t)$ must satisfy the following differential equation

$$\frac{d\Sigma(t)}{dt} = \sigma_F^2 - \sigma(t)^2. \quad (28)$$

Define $d\bar{Z}(t) := \Sigma(t)^{-1} \cdot \sigma(t) \cdot \sigma_F \cdot [B(t) - \bar{B}(t)] \cdot dt + dZ(t)$. The process $\bar{Z}(t)$ is a standardized Brownian motion under the market's filtration, because $\bar{B}(t)$ is the market's best estimate of $B(t)$ at time t . Equation (25) can then be written in the simplified form

$$d\bar{B}(t) = \sigma(t) \cdot \sigma_F^{-1} \cdot d\bar{Z}(t). \quad (29)$$

Substituting $d\bar{Z}(t)$ for $d\bar{B}(t)$ in equation (27), we write the equilibrium price process as the martingale

$$\frac{dP(t)}{P(t)} = \sigma(t) \cdot d\bar{Z}(t). \quad (30)$$

Thus, $\sigma(t)$ indeed measures returns volatility, consistent with our notation σ throughout this paper. Returns volatility $\sigma(t)$ measures the rate at which new information is being incorporated into prices. According to equation (26), returns volatility $\sigma(t)$ is stochastic, even though fundamental volatility σ_F is constant.

Equations (28) and (30) have a simple intuition. When returns volatility $\sigma(t)$ is greater than fundamental volatility σ_F , new information is being incorporated into prices faster than new fundamental uncertainty is unfolding, and the difference $\sigma_F^2 - \sigma^2(t)$ represents the rate at which pricing error $\Sigma(t)$ is decreasing. When returns volatility $\sigma(t)$ is smaller than fundamental volatility σ_F , new fundamental uncertainty is unfolding faster than information is being incorporated into prices, and the difference $\sigma_F^2 - \sigma^2(t)$ represents the rate at which pricing error $\Sigma(t)$ is increasing.

When $\Sigma'(t) = 0$, we shall say that $\Sigma(t)$ has reached a “conditional steady state” in which the unfolding of new fundamental volatility and the incorporation of new private information into prices are in balance, i.e., $\sigma(t) = \sigma_F$. We use this terminology because the right-hand side of equation (28) does not converge to a constant. Indeed, the value of $\gamma(t)$, which follows a diffusion, is constantly changing and this makes $\sigma(t)$ change as well. If $\gamma(t)$ were to remain constant for a long period of time, the value of $\Sigma(t)$ would converge to the conditional steady state given by

$$\Sigma(t) = \frac{\sigma_F^2}{\gamma(t) \cdot \theta^2 \cdot \tau}, \quad (31)$$

obtained from the definition of $\sigma(t)$ in equation (26) and equation (28) when $\Sigma'(t) = 0$.

As changes in prices $P(t)$ lead to immediate changes in market capitalization and therefore changes in the arrival rate of bets $\gamma(t)$, the value of $\Sigma(t)$ gradually drifts towards a “conditional steady state” value. While it is the *level* of $\gamma(t)$ that follows a diffusion, it is the *derivative* of $\Sigma(t)$ that follows a diffusion. Therefore, as $\gamma(t)$ changes, the value of $\Sigma(t)$ moves smoothly towards a conditional steady-state value which it is constantly chasing.

According to equation (31), more accurate signals and more frequent bets make market prices more accurate. The invariance theorem discussed next applies both in the conditional steady state and outside it.

The structural model also implies a particular relationship between pricing accuracy and resiliency. Recall that resiliency $\rho(t)$ denote the mean reversion rate at which pricing errors resulted from execution of uninformative bets disappear, as prices are pushed back towards fundamentals whenever an informative bet is execute. Equations (25) and (26) imply that the difference $B(t) - \bar{B}(t)$ follows the mean-reverting process²

$$d[B(t) - \bar{B}(t)] = -\frac{\sigma(t)^2}{\Sigma(t)} \cdot [B(t) - \bar{B}(t)] \cdot dt + dB(t) - \frac{\sigma(t)}{\sigma_F} \cdot dZ(t). \quad (32)$$

²Black (1986) contains the intuition that because transitory noise affects prices, prices have twice the returns variance as fundamental value, and this is associated with mean reversion in returns. Equation (32) shows that, in a steady state, Black’s intuition applies to the log-ratio of prices to fundamental value, not to prices themselves (equation (30)), which have a martingale property. Black (1986) may have been confusing the properties of prices with the properties of the ratio of fundamental value to price.

Thus, we have

$$\rho(t) = \frac{\sigma(t)^2}{\Sigma(t)}. \quad (33)$$

Holding returns volatility constant, resiliency is greater in markets with higher pricing accuracy.

This equation (33) suggests an empirical strategy for calibrating Black’s measure of market efficiency, which is difficult to observe directly. A value for Σ can be inferred indirectly from an estimate of resiliency ρ , which can be obtained by examining how fast price effects from noise trades die out over time (or from examining how long traders hold actively managed positions based on a more general model in which positions are liquidated after some fraction of their information content is revealed in prices). As discussed before, for example, if a stock’s annual volatility is about 35% and $\Sigma^{1/2} = \ln(2)/1.64$, then prices are about 1.50 years “behind” fundamental value, i.e., $\Sigma/\sigma^2 = (\ln(2)/1.64)^2/0.35^2 \approx 1.50$. The error $B(t) - \bar{B}(t)$ in equation (32) mean-reverts to zero at rate $\rho = \sigma^2 \cdot \Sigma^{-1} = 0.35^2/(\ln(2)/1.64)^2 = 0.69$ per year. This implies that the half-life of the price impact of a noise trade is equal to $\ln(2)/\rho \approx 1$ year. Black (1986) could, therefore, have equivalently defined an efficient market as one in which the half-life of the price impact of noise trades is less than one year.

The endogenous quantities in the model are all functions of the two “state variables” $P(t)$ and $\Sigma(t)$, which evolve stochastically according to equations (28) and (30), with stochastic returns volatility given by equation (26). These equations describe interesting dynamics. When trading volume is high, bets arrive quickly and $\Sigma(t)$ moves quickly towards its conditional steady state level; returns volatility remains close to fundamental volatility; $\Sigma(t)$ does not deviate far from its conditional steady state level. When trading volume is low, bets arrive slowly and $\Sigma(t)$ adjusts slowly towards its conditional steady state level; returns volatility may remain below fundamental volatility for extended periods of time.³

2 Solution of the Model with Invariance

The following invariance theorem states that this structural model implies, under the assumption of constant cost of generating an informed signal, all of the invariance hypotheses in Kyle and Obizhaeva (2016) and reveals specific connections among the hypotheses.

Invariance Theorem. *Assume the cost c_I of generating a signal is an invariant constant and let $m := E\{|\tilde{i}(t)|\}$ define an additional invariant constant. Then, the*

³It can be shown that (with probability one) (1) a stock’s fundamental value will eventually become very small (since it follows a geometric Brownian motion which is also a martingale), (2) both the bet arrival rate and returns volatility will eventually converge to zero, (3) and $\Sigma(t)$ will eventually become unboundedly large. This is consistent with the interpretation that almost all stocks are eventually de-listed. As Keynes would say, in the long run, all companies are dead.

invariance conjectures hold: The dollar risk transferred by a bet per unit of business time $\tilde{I}(t)$ is a random variable with an invariant distribution $\bar{c}_B \cdot \tilde{i}(t)$ and the expected cost of executing a bet $\bar{C}_B(t)$ is an invariant constant \bar{c}_B .⁴

$$\tilde{I}(t) := \frac{P(t) \cdot \tilde{Q}(t) \cdot \sigma(t)}{\gamma(t)^{1/2}} = \frac{\tilde{Q}(t)}{V(t)} \cdot W(t)^{2/3} \cdot (m \cdot \bar{c}_B)^{1/3} = \bar{c}_B \cdot \tilde{i}(t). \quad (34)$$

$$\bar{C}_B(t) = \bar{c}_B := c_I \cdot \theta / (1 - \theta). \quad (35)$$

The expected dollar transaction cost of executing a bet of size Q is the same function of the dollar size of risk it transfers in units of business time I :

$$C_B(Q) := \frac{1}{\bar{c}_B} \cdot I^2, \quad \text{where} \quad I(t) \equiv \frac{P(t) \cdot Q(t) \cdot \sigma(t)}{\gamma(t)^{1/2}}. \quad (36)$$

The number of bets $\gamma(t)$, their size $\tilde{Q}(t)$, liquidity $L(t)$, pricing accuracy $1/\Sigma(t)^{1/2}$, and market resiliency $\rho(t)$ are related to price $P(t)$, share volume $V(t)$, volatility $\sigma(t)$, and trading activity $W(t) = P(t) \cdot V(t) \cdot \sigma(t)$ by the following invariance relationships:

$$\gamma(t) = \left(\frac{\lambda(t) \cdot V(t)}{\sigma(t) \cdot P(t) \cdot m} \right)^2 = \left(\frac{E\{|\tilde{Q}(t)|\}}{V(t)} \right)^{-1} = \frac{(\sigma(t) \cdot L(t))^2}{m^2} = \frac{\sigma(t)^2}{\theta^2 \cdot \tau \cdot \Sigma(t)} = \frac{\rho(t)}{\theta^2 \cdot \tau} = \left(\frac{W(t)}{m \cdot \bar{c}_B} \right)^{2/3}. \quad (37)$$

Here, τ is the precision of a signal, and θ is the fraction of information $\tilde{i}(t)$ incorporated by an informed trade. The price follows a martingale with stochastic returns volatility $\sigma(t) := \theta \cdot \tau^{1/2} \cdot \Sigma(t)^{1/2} \cdot \gamma(t)^{1/2}$. The arrival rate of bets $\gamma(t)$ sets the pace of business time.

These equations directly correspond to empirical hypothesis in Kyle and Obizhaeva (2016). The bet size invariance says that the distribution of the dollar risk transferred by a bet is the same when the dollar risk is measured in units of business time; it directly corresponds to equation (34). The transaction costs invariance says that the expected dollar transaction cost of executing a bet is the same function of the size of the bet when the bets size is measured as the dollar risk it transfers in units of business time; it directly corresponds to equation (36). Equation (35) is the unconditional version of this statement.

Equation (37) summarizes empirical implications about bet arrival rate, bet size as well as price impact and adds new empirical implications about pricing accuracy and resiliency.

⁴In this theorem, “invariance” means that the values of the constants c_I , m , and $\bar{C}_B(t)$ as well as the distribution of $\tilde{I}(t)$ do not vary with time. In a model with different securities, they would not vary across securities either.

Proof of Invariance Conjectures and Relationships. Using equations (17), (20), and (21), we derive equation (35). The value $\bar{c}_B = \theta \cdot c_I / (1 - \theta)$ is constant across stocks, since c_I is constant across stocks by assumption and θ is equal by proof to $\theta = 1/2$ for risk-neutral informed traders (or, alternatively, constant by assumption for more general cases).

Dividing the definition $\tilde{I}(t) := P(t) \cdot \tilde{Q}(t) \cdot \sigma(t) \cdot \gamma(t)^{-1/2}$ by the equation $\bar{C}_B(t) := \lambda(t) \cdot E\{\tilde{Q}(t)^2\}$, plugging in the definitions of $\tilde{Q}(t)$ and $\Delta\tilde{B}_I(t)$ from equations (5) and (9), and using definition (26), we obtain the third equality $\tilde{I}(t) = \bar{c}_B \cdot \tilde{i}(t)$ in equation (34); the first equality in equation (34) is the definition of $\tilde{I}(t)$ and the second equality (involving $W^{2/3}$) will follow from equation (37) below. Since $\tilde{i}(t)$ is invariant by assumption and \bar{c}_B is invariant by proof, $\tilde{I}(t) = \bar{c}_B \cdot \tilde{i}(t)$ has an invariant distribution. Note that both $\tilde{I}(t)$ and \bar{c}_B are measured in dollars while $\tilde{i}(t)$ represents unit-less information.

To derive equation (36), we get $\lambda(t)$ from the equation $\bar{c}_B := \lambda(t) \cdot E\{\tilde{Q}(t)^2\}$, plug it into $C_B(\tilde{Q}(t)) = \lambda(t) \cdot \tilde{Q}(t)^2$, then plug $\tilde{Q}(t)$ obtained from the definition $\tilde{I}(t) := P(t) \cdot \tilde{Q}(t) \cdot \sigma(t) \cdot \gamma(t)^{-1/2}$, and finally use $\tilde{I}(t) = \bar{c}_B \cdot \tilde{i}(t)$ from equation (34) and $E\{\tilde{i}^2\} = 1$.

To derive equation (37), we first summarize all necessary information. We write equation (14) for daily volume, equation (19) for expected costs using equation (9), and equation (26) for returns volatility using equations (5) and (9) as well as $E\{\tilde{i}^2(t)\} = 1$ as a system of three equations:

$$\gamma(t) \cdot E\{|\tilde{Q}(t)|\} = V(t), \quad (38)$$

$$\bar{c}_B = \lambda(t) \cdot E\{\tilde{Q}(t)^2\}, \quad (39)$$

$$\gamma(t) \cdot \lambda(t)^2 \cdot E\{\tilde{Q}(t)^2\} = P(t)^2 \cdot \sigma(t)^2. \quad (40)$$

These three equations reflect the economic assumptions important for deriving invariance. Equation (38) says that observable trading volume results from bets. Equation (39) says that the expected price impact cost of a bet is \bar{c}_B (which is constant from equation (35) that we have proved) assuming the bet has linear price impact costs. Equation (40) says that the price impact of a bet generates returns volatility, also under the assumption that price impact is linear.⁵

In the three equations (38), (39), and (40), one can think of $\gamma(t)$, $\lambda(t)$, $E\{\tilde{Q}(t)^2\}$, and $E\{|\tilde{Q}(t)|\}$ as unknown variables to be solved for in terms of known variables $V(t)$, \bar{c}_B , $P(t)$, and $\sigma(t)$. Since there are three equations and four unknowns, we need a fourth equation. The fourth equation is the moment ratio

$$m = \frac{E\{|\tilde{Q}(t)|\}}{[E\{\tilde{Q}(t)^2\}]^{1/2}}. \quad (41)$$

⁵Note that the concepts of “price impact cost” and “price impact,” which both depend on the linear price impact parameter $\lambda(t)$ in the model, actually represent somewhat different concepts, which happen to be the same in a dealer market model in which markets are semi-strong efficient and dealers make zero profits based on linear signal processing. In the structural model, it is a derived result that price impact and transaction costs are linear.

Since $m := E\{|\tilde{i}|\}$, this equation follows from the fact that $\tilde{Q}(t)$ is a linear multiple of $\tilde{i}(t)$ and $\tilde{i}(t)$ has a variance of one. Since $\tilde{i}(t)$ is assumed to have a normal distribution with a variance of one, we have $m = (2/\pi)^{1/2}$. For a different distributional assumptions about information, m would have a different value. If we think of m as an exogenous parameter, we now have four equations (38)-(41) in four unknowns.

Using the definition of m and the definition of trading activity $W(t) = P(t) \cdot V(t) \cdot \sigma(t)$, we can solve equations (38), (39), and (40) for $\gamma(t)$, $E\{|\tilde{Q}(t)|\}$, and $\lambda(t)$, as follows. Multiply the product of (38) and (39) by the square root of (40) and solve for $\gamma(t)$ to obtain

$$\gamma(t) = (m \cdot \bar{c}_B)^{-2/3} \cdot W(t)^{2/3}. \quad (42)$$

Divide the product of (40) and the square of (39) by (38) and solve for $E\{|\tilde{Q}(t)|\}$ using (41) to obtain

$$E\{|\tilde{Q}(t)|\} = (m \cdot \bar{c}_B)^{2/3} \cdot V(t) \cdot W(t)^{-2/3}. \quad (43)$$

Divide the product of (40) and the square root of (39) by (38) and solve for $\lambda(t)$ to obtain

$$\lambda(t) = \left(\frac{m^2}{\bar{c}_B}\right)^{1/3} \cdot \frac{1}{V(t)^2} \cdot W(t)^{4/3}. \quad (44)$$

Equation (43) and the definition of illiquidity $1/L(t) := \bar{c}_B/[P(t) \cdot V(t)]$ imply that $1/L$ satisfies

$$\frac{1}{L(t)} = \left(\frac{m^2}{\bar{c}_B}\right)^{-1/3} \cdot \sigma(t) \cdot W(t)^{-1/3}. \quad (45)$$

Equations (26), (33), and (42) imply that pricing accuracy $1/\Sigma(t)^{1/2}$ and resiliency $\rho(t)$ satisfy

$$\rho(t) = \frac{\sigma(t)^2}{\Sigma(t)} = \left(\frac{1}{m \cdot \bar{c}_B}\right)^{2/3} \cdot \theta^2 \cdot \tau \cdot W(t)^{2/3}. \quad (46)$$

Equations (42) for $\gamma(t)$, (43) for $\tilde{Q}(t)$, (45) for $1/L(t)$, and (46) for $1/\Sigma(t)^{1/2}$ and $\rho(t)$ are summarized in equation (37). This completes the proof of the theorem.

Link to Empirical Hypotheses. Equations (42) for $\gamma(t)$, (43) for $\tilde{Q}(t)$, (45) for $1/L(t)$ are respectively equivalent to equations (7), (8), and (15) implied by the market microstructure invariance hypotheses, since $E\{|\tilde{i}|\} = m$ implies

$$E\{|\tilde{I}|\} = m \cdot \bar{c}_B \quad (47)$$

from (34). The predictions (46) for $1/\Sigma(t)^{1/2}$ and $\rho(t)$ are new.

Since trading activity $W(t)$ and its components are observable, we can empirically infer values of $\gamma(t)$, $E\{|\tilde{Q}(t)|\}$, $1/L(t)$, $\lambda(t)$, $1/\Sigma(t)^{1/2}$ and $\rho(t)$ from equation (37) if the values of the three constants m , \bar{c}_B , and $\theta^2\tau$ are known. If these constants

are not known, then they can be estimated as intercepts in regressions of logs of the corresponding variables on logs of trading activity.

Invariance relationships come about through the following intuition: Suppose the number of noise traders increases for some exogenous reason. In the structural model, this happens when the share price and therefore market capitalization increases, keeping the share turnover of noise traders constant. To be specific, assume that the number of noise traders increases by a factor of 4. As a result, market depth increases and, consequently, the number of informed traders increases, since their bets now are more profitable. The structural model shows that the number of informed traders eventually increases by a factor of 4 as well, and each of their bets accounts for a 4 times smaller fraction of returns variance. Returns volatility per unit of business time decreases by a factor of 2 (the square root of 4). The structural model shows that pricing accuracy and liquidity both increase by a factor of 2, as a result of which informed traders exactly cover the cost of private signals by submitting bets 2 times as large as before. Overall dollar volume in the market increases by a factor of 8. Thus, the “one-third, two-thirds” intuition comes about: One-third of the increase in dollar volume comes from changes in bet size ($8^{1/3} = 2$) and two-thirds comes from changes in the number of bets ($8^{2/3} = 4$).

3 Discussion

The model adds additional structure that imposes restrictions on the empirical invariance hypotheses outlined in Kyle and Obizhaeva (2016). These additional assumptions impose a particular structure on the proportionality constants in invariance relationships and thus allow us to write these disconnected relationships in a consolidated form of the invariance theorem (37).

The level of trading activity $W(t)$ and its components—prices $P(t)$, share volume $V(t)$, and returns volatility $\sigma(t)$ —are “macroscopic” quantities whose value can be estimated from aggregate market data, e.g., from the CRSP dataset. The bet arrival rate $\gamma(t)$, the bet size $\tilde{Q}(t)$, the average cost of a bet $1/L(t)$, pricing accuracy $\Sigma(t)^{1/2}$, and resiliency $\rho(t)$ are granular “microscopic” quantities whose values are difficult to observe even with data on individual trades by individual traders. Since knowledge of the constants \bar{c}_B , m , and $\tau \cdot \theta^2$ makes it possible to infer microscopic quantities from macroscopic quantities using equation (37), these constants play roles in our structural model somewhat similar to the role played by Boltzmann’s constant or Avogadro’s number in physics.

For example, the structural model implies a particular relationship between the invariance of bet sizes and transaction costs. Equation (34) suggests that the bet size invariant \tilde{I} and transaction costs invariant \bar{c}_B satisfy the restriction $E|\tilde{I}| = m \cdot \bar{c}_B$ or, equivalently, that the standard deviation of \tilde{I} is equal to \bar{c}_B . This restriction follows from the assumption that market makers break even.

In this paper, invariance relationships for bets and transaction costs were derived

based on the assumption that the cost of executing a bet \bar{c}_B is constant across stocks. The structural model shows that \bar{c}_B is not the “deepest” structural parameter in the model. The result $\bar{c}_B = \theta \cdot c_I / (1 - \theta)$ —for example, if $\theta = 1/2$ then $\bar{c}_B = c_I$ —implies that \bar{c}_B is constant across stocks if the “deeper” structural parameters c_I and θ are constant across stocks. It is useful to think of the cost of private information c_I as proportional to the average wages of finance professionals, adjusted for their productivity or effort required to generate one bet. The productivity-adjusted wage of a finance professional is therefore a “deeper” parameter than the endogenous cost of executing a bet \bar{c}_B . The invariance relationships in equation (37) result from finance professionals allocating their skills across different markets to maximize the value of trading on the signals they generate. The assumption that distinct bets result from distinct pieces of private information implies a particular level of granularity for both signals and bets.

It is interesting that invariance relationships relating the granularity of bets to their costs depend on the first absolute moment of the distribution of signals m , but not on the precision of signals τ .

The structural model also shows that the invariance of pricing accuracy and resiliency requires stronger assumptions: In addition to \bar{c}_B and m being constant, the informativeness of a bet $\tau \cdot \theta^2$, measured as the product of signal precision τ and the squared fraction of informed traders θ^2 , must be constant across time (or across stocks) as well. Based on equation (37), the principles of *pricing accuracy invariance* and *resiliency invariance* say that pricing accuracy is the same if its reciprocal is scaled by returns volatility per unit of business time, and market resiliency is the same if it is measured in units of business time.

Although the structural model is motivated by the time series properties of a single stock as its market capitalization changes, the model applies cross-sectionally across different securities under the assumption that the exogenously assumed cost of a private signal c_I , the shape of the distribution of signals m , and the informativeness of bets $\tau \cdot \theta^2$ are constant across all markets.

Robustness of Assumptions. Our structural model makes restrictive assumptions which lead to specific properties of equilibrium. Private signals are normally distributed. Bets are linear functions of private signals and are therefore also normally distributed. Price impact is linear in bet size. Informed traders place one bet in a dealer market setting rather than shredding bets into many pieces and executing them at an equilibrium speed over time. Dealers do not make profits; there are no bid-ask spread costs; there are no “effective bid-ask spread” costs related to immediate reversal of temporary price impact in executing bets.

The empirical results are not consistent with the “linear-normal” features of the structural model, but are consistent with the more general empirical hypotheses concerning bet size and transactions costs. Kyle and Obizhaeva (2016) find that the size of unsigned bets closely fits a log-normal distribution, not a normal distribution. Also,

a linear price impact model predicts transaction costs reasonably well, but a square root model predicts transaction costs better than a linear model. We conjecture that it should be possible to modify our structural model to accommodate non-normally distributed bet size, non-linear price impact, and dynamic execution of bets at an equilibrium speed proportional to the rate at which business time unfolds.

Although the structural model assumes that fundamental volatility σ_F , shares outstanding N , and noise-trader turnover rate η are constants, it is also straightforward to modify the model so that these quantities vary over time or across stocks.

Comparison with Kyle (1985). As our model, the continuous-time model of Kyle (1985) is a dynamic model of trading between a risk-neutral monopolistic informed trader, noise traders, and competitive risk-neutral market makers. There are, however, several modelling differences. First, the fundamental value does not change over time and realizes once at the end of the game at time T . Second, there is only one informed trader who gets a private signal at the beginning and trades on it gradually throughout the game.

The strategies of informed traders are similar in both models. In Kyle (1985), the informed trader follows a strategy $dx(t) = \beta(t) \cdot (\tilde{v} - p(t)) \cdot dt$; the intensity coefficient $\beta(t)$ is not determined from the optimization problem of the informed trader, and any function $\beta(t)$ is optimal as long as the strategy is smooth. In our model, informed traders trade $\tilde{Q}(t) = \theta \cdot \lambda(t)^{-1} \cdot E\{F(t) - P(t) \mid \Delta \bar{B}_I(t)\}$ from equation (9); the intensity coefficient $\theta \cdot \lambda(t)^{-1}$ is also left undetermined in the optimization problem of informed traders, only $\theta = 1/2$ is obtained.

The price impact parameters $\lambda(t)$ are not identified from the market efficiency conditions (15) in both models. Instead, the market depth is pinned down by the condition saying that all volatility results from trading. In Kyle (1985), the price impact $\lambda = \sigma_v / \sigma_u$ is obtained from the restriction $\Sigma(T) = 0$ saying that the error variance disappears by the end of the game, as trading of the informed trader pushes prices towards fundamentals. In our model, the price impact $\lambda = (P(t)\sigma(t)) / (\gamma(t) \cdot E\{\tilde{Q}(t)^2\})^{1/2}$ is obtained from the similar equation (40), where $(\gamma(t) \cdot E\{\tilde{Q}(t)^2\})^{1/2}$ is a stochastic proxy for the standard deviation of order imbalances.

Finally, the equilibrium price process is a martingale in both papers. The difference is that price volatility is a constant equal to variance of fundamentals σ_v in Kyle (1985), whereas price volatility $\sigma(t)$ is stochastic in our model.

4 Conclusion

The empirical invariance hypotheses of Kyle and Obizhaeva (2016) are likely to hold more generally than under the somewhat restrictive assumptions of our structural model. The structural model is to be interpreted as a “proof of concept,” while the invariance hypotheses should apply more generally.

To derive invariance relationships, we mostly relied on the four equations (38)-(41). These equations capture generic properties of trading. The invariance predictions ultimately rely on several general assumptions that (1) order flow creates volume and induces volatility and (2) ratio of moments of bet size distributions is stable across assets and time. We therefore conjecture that invariance relationships can be obtained in the context of other market microstructure models as well.

References

- Andersen, Torben G., Oleg Bondarenko, Albert S. Kyle, and Anna A. Obizhaeva.** 2015. "Intraday Trading Invariance in the E-Mini S&P 500 Futures Market." Working Paper, available at <http://dx.doi.org/10.2139/ssrn.2693810>.
- Black, Fischer.** 1986. "Noise." *The Journal of Finance*, 41: 529–543.
- Clark, Peter.** 1973. "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices." *Econometrica*, 41: 135–155.
- Fama, Eugene.** 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance*, 25(2): 383–417.
- Glosten, Lawrence R., and Paul R. Milgrom.** 1985. "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders." *The Journal of Financial Economics*, 14: 71–100.
- Harris, Lawrence.** 1987. "Transaction Data Tests of the Mixture of Distributions Hypothesis." *Journal of Financial and Quantitative Analysis*, 22(02): 127–141.
- Karpoff, Jonathan M.** 1987. "The Relation between Price Changes and Trading Volume: A Survey." *Journal of Financial and Quantitative Analysis*, 22(1): 109–126.
- Kyle, Albert S.** 1985. "Continuous Auctions and Insider Trading." *Econometrica*, 53(6): 1315–1335.
- Kyle, Albert S., and Anna A. Obizhaeva.** 2016. "Market Microstructure Invariance: Empirical Hypotheses." *Econometrica*, accepted for publication, available at <http://dx.doi.org/10.2139/ssrn.2722524>.
- LeRoy, Stephen F.** 1989. "Efficient Capital Markets and Martingales." *Journal of Economic Literature*, 27(4): 1583–1621.
- Tauchen, George E., and Mark Pitts.** 1983. "The Price Variability-Volume Relationship on Speculative Markets." *Econometrica*, 51: 485–505.

Treynor, Jack. 1995. "The Only Game in Town." *The Financial Analysts Journal*, 51(1): 81–83 (reprinted from *The Financial Analysts Journal* 22, 1971, 12–14, 22).