

Adverse Selection and Liquidity: From Theory to Practice

Albert S. Kyle

Anna A. Obizhaeva*

June 3, 2018

Abstract

This paper shows how to map predictions of theoretical models of market microstructure into operational empirical measures of liquidity. A meta-model implies an empirical measure of liquidity, denoted L , which describes various characteristics of trading and funding liquidity such as trading costs, bet sizes, haircuts, and capital requirements. When mapped into existing models of adverse selection, the meta-model also describes precisely how adverse selection shows up in pricing accuracy and resiliency. The meta-model is consistent with models of both block trading and flow trading. It highlights a deep connection between time and adverse selection.

JEL: G10, G12, G14, G20

Keywords: market microstructure, invariance, liquidity, adverse selection, market impact, bid-ask spread, bet size, market efficiency, dimensional analysis, leverage neutrality.

*Kyle: University of Maryland, College Park, MD 20742, USA, akyle@rhsmith.umd.edu. Obizhaeva: New Economic School, Moscow, Skolkovo, 143026, Russia, aobizhaeva@nes.ru. Albert S. Kyle has worked as a consultant for various companies, exchanges, and government agencies. He has worked for the U.S. Department of Justice as a consultant on issues related to competition in financial markets. He is a non-executive director of a U.S.-based asset management company.

Practitioners and academics have studied market liquidity for a long time; it is an elusive concept. Despite voluminous amounts of data and many theoretical papers, it has proven difficult to measure liquidity and to map theoretical concepts like adverse selection into operational empirical predictions about prices and quantities. This paper first uses a meta-model to derive a simple empirical measure of liquidity, maps theoretical models of market microstructure into the meta-model, and uses the mapping to show how adverse selection in financial markets is related to market liquidity.

Liquid financial markets are dominated by institutional investors who place “bets” by trading significant quantities, either as large discrete blocks or as many smaller trades spread out over time. In more liquid markets, both the size and number of bets are larger. Each asset has its own “business time,” which measures the speed with which bets arrive. Execution of bets move prices. Since some bets are based on private information, the resulting price impact incorporate information into prices. Large traders with private information constrain the sizes of their bets to control market impact costs. Meaningful models of market microstructure are therefore inherently noncompetitive, and market power may restrict the trading of the most significant asset managers much more than risk aversion or assets under management.

Market microstructure research is different from asset pricing research. Asset pricing is largely concerned with how risk is priced in competitive markets under the assumption that asset prices reveal information about cash flows while expected returns also fluctuate. Market microstructure is concerned with how market liquidity is related to the way in which private information finds its way into prices through the process of trading in imperfectly competitive markets.

Our methodology begins by dividing microstructure quantities into three categories based on ease of observability and variation across assets. The first category includes quantities which are easy to observe or estimate, such as price, volume, and return volatility. The second category includes quantities which are both hard to observe and vary greatly across assets. This category includes bet sizes, bet arrival rates, market depth, and parameters defining market impact functions. Although researchers are interested in understanding these quantities, measuring them has proven difficult because required data are often proprietary. Understanding how these quantities vary across assets and across time is key to understanding market microstructure. Our methodology emphasizes a third category, defined as “invariant” quantities which are hard to observe but can be hypothesized to vary little, if at all, across assets. These quantities include the dollar market impact cost of a bet, moments measuring the shape but not the scaling of bet size, the precision of informed traders’ signals, and the fraction of traders who are informed.

“Market microstructure invariance” is the hypothesis that the third category of quantities do

not vary across assets or time; they are like deep parameters in a structural model. The strategy of this paper is to use the invariance of parameters in the third category to derive universal scaling laws describing the behavior of hard-to-observe variables. To make the strategy as broadly applicable as possible, we derive the scaling laws from a “meta-model.” The meta-model is a set of reduced-form equations which describe properties that many different models of market microstructure should share. The equations involve all three types of quantities: easy to observe, highly variable but hard to observe, and invariant ones. Implementing the strategy successfully requires solving for the highly variable but hard-to-observe quantities in terms of easy-to-observe quantities and invariant quantities. Essentially, this requires having as many equations in the meta-model as there are hard-to-observe quantities.

The meta-model begins with three simple equations which assert that (1) market volume is the sum of all bets, (2) bets have market impact which is a function of their size, and (3) return volatility results from the cumulative, random price impact of bets. These equations are a meta-model in the sense that they are empirically sensible implications of many theoretical market microstructure models. To make it possible to solve the model, these equations are supplemented with equations related to invariant quantities. The most important is the bet-cost invariance equation, which asserts that the dollar expected market impact cost of executing a bet is an invariant constant C . In addition, there are two equations asserting the invariance of various moment ratios describing the shape of the distribution of bet sizes. The result is six equations, which can be solved for six difficult-to-observe quantities that vary greatly across assets. The solution to the six equations defines universal scaling laws for market microstructure, including a universal market impact function.

The solution to the meta-model can be naturally described using a liquidity measure which we call L . The value of L is proportional to the cube root of the ratio of dollar volume to returns variance, with the proportionality constant defined by invariant quantities. The quantity L is a universal measure of market liquidity, which can be used to describe bet size, the number of bets, and the price impact function. The value of L is scaled so that $1/L$ measures the average dollar-weighted price-impact cost of executing a bet as a fraction of the value traded.

The results implied by the meta-model can also be derived in two other universal ways. First, following Kyle and Obizhaeva (2016a), the meta-model can be derived directly from invariance principles related to the concept of “business time,” defined as the rate at which bets arrive into a market. The speed of business time, which we denote γ , is proportional to the product of returns variance and L^2 . Second, following Kyle and Obizhaeva (2017a), the implications of the meta-model can be derived from dimensional analysis and leverage neutrality using an approach similar to that of physicists. “Leverage neutrality,” the hypothesis that cash can be exchanged freely, is a market-microstructure version of the Modigliani–Miller theorem.

Liquidity L and business time γ and are two different ways of measuring liquidity. The cost of transferring a given dollar value, measured as a fraction of the value traded, is related to $1/L$. The dollar cost of transferring a risk of a given size is related to γ . For example, a given dollar quantity of 30-day Treasury bills may be more liquid than the same dollar quantity of S&P 500 futures contracts when measured using $1/L$ but less liquid when measured using γ .

The six meta-model equations define market microstructure as relationships describing how prices, volumes, and volatilities relate to bet sizes, number of bets, and price impact. Conspicuously missing from the meta-model is any sense of underlying economic fundamentals related to adverse selection, the information content of bets, or the accuracy with which prices measure fundamentals. A researcher implementing the meta-model model with perfect empirical precision would understand how quantities traded result in price changes but would have no idea how to interpret the information content of these price changes. The apparent lack of economics in the meta-model equations mimics the apparent lack of economics traders themselves see in a market. Instead of seeing information arriving as the economic laws of supply and demand unfold, traders in a market see a plethora of seemingly random orders, bids, offers, trade executions, and price fluctuations. Indeed, many successful market makers have little understanding of the economic forces governing prices in the markets where they make a good living by trading.

A deeper economic interpretation of the solution of the meta-model is obtained by mapping the meta-model into four theoretical economic models of market microstructure. The four models show how to relate the meta-models measures of liquidity to the economic forces of adverse selection and information flow in financial markets. The theoretical models link the measure L , in empirically implementable manner, to the error variance of prices relative to fundamentals and to market resiliency. "Market resiliency" is defined as the rate at which the flow of new information into the market makes past pricing errors, related perhaps to noise trading, die out over time.

The first three models are all models of block trading in that traders place bets by trading given quantities all at once. While block trading was common in the 20th century, traders of U.S. equities in the 21st century place bets by breaking blocks into tiny piece and executing many small orders over time. This shift from block trading to flow trading has economic underpinnings related to how changes in technology and regulation have affected optimal exercise of market power.

First, we map the meta-model into the theory of Treynor (1995). This makes it possible to interpret the bid-ask spread as information about fundamental value and the probability of a trade being informed. The invariance assumption connects the width of the bid-ask spread to the size of bets in the model in an empirically realistic manner.

Second, we map the meta-model into the one-period model of Kyle (1985). This makes it possible to show why market impact may be a linear function of the size of the bet. Market depth reflects the informativeness of the order flow and the size of trades by noise traders. The invariance assumption provides a universal formula for implementing the market depth parameter λ empirically. It also connects the invariance assumption to the cost of information.

Third, we add adverse selection to the meta-model by describing an informal dynamic model of informed and noise trading related to the model of Kyle and Obizhaeva (2017*b*).

Fourth, to understand these changes, we map the meta-model into a model of smooth trading with elements similar to the smooth trading model of Kyle, Obizhaeva and Wang (2018) and the dynamic model. The smooth trading model captures recent changes in the way traders trade. The previous three models are models of “block trading” in which traders enter the market once, execute a block trade of given size once, and then do not trade again. The block trading idea is consistent with the way dealer markets, including the upstairs market for stocks at the NYSE, functioned in the 1990s. As a result of changed order handling rules, improvements in electronic trading technology, regulations which fragmented trading venues (such as Regulation NMS in the United States and MiFiD in the European Union), and reduced minimum tick size, traders began to trade gradually through time, shredding their bets into tiny pieces, frequently using Volume-Weighted-Average-Price (VWAP) trades. Empirically, average trade size in Trades and Quotations (TAQ) data has fallen greatly in recent years (see Kyle, Obizhaeva and Tuzun (2017)).

In the smooth trading model, there is both permanent price impact and temporary price impact. Permanent price impact is measured by the parameter λ as before. Temporary price impact, denoted κ , measures the effect of the time derivative of a trader’s inventory on prices. When a trader stops trading, temporary price impact immediately disappears. In the smooth trading model, trading a large block over a short period of time triggers large temporary price movements, which reverse after the trade is completed. In equilibrium, the speed with which traders trade and the instantaneous impact of their trades are related to the resiliency of prices. The resiliency parameter ρ maps temporary price impact into permanent price impact and also maps flow trading into intended bets. More liquid markets are more resilient and therefore allow traders to trade more quickly. The smooth trading model shows that the speed of trading and the resilience of prices map directly into the rate at which private information is flowing into the market through informed trading. This information flow is precisely consistent with the way Black (1986) defines market efficiency as the accuracy of prices as estimates of fundamental value. It is also consistent with the Fama’s definition of the efficient markets hypothesis as market prices reflecting publicly available information.

To summarize, the mapping of these four models into the meta-model shows how private

information is revealed in prices through trading based on adverse selection, even though adverse selection is not present in the meta-model itself. The mappings show a universal, deep connection between adverse selection and the meta-model.

Practitioners often define liquidity as the ability to sell an asset quickly without significant price impact. This practitioner definition incorporates time into the definition of liquidity in a manner precisely consistent with the two dynamic models, especially the model of smooth trading. The meta-model describes trading liquidity. Economists often make a distinction between trading liquidity and funding liquidity. Our analysis shows that trading liquidity and funding liquidity are closely related through the concept of time. Trading liquidity and funding liquidity are essentially the same concepts, connected through the concept of time; both are described by the liquidity measures L and γ . Optimal repo haircuts, capital requirements, and recapitalization costs are all proportional to $1/L$. The connection occurs because selling defaulted assets, liquidating a failed bank, and raising new capital are all important funding activities which take a period of time related to the trading liquidity of the relevant assets themselves. More information flow makes assets more liquid.

Empirical Evidence. Although this paper is theoretical, it is meant to provide a roadmap for empirical research in market microstructure. The meta-model predicts scaling laws with exponents of one-third or two-thirds for microstructure variables. For the empirical approach to work, it is necessary to find evidence of these scaling laws in data. Our research so far generally finds that these scaling laws hold as economically close approximations, but not exactly. Here are some examples.

Kyle and Obizhaeva (2016*a*) document invariance-implied relationships for the size of meta-orders and transaction costs for a sample of portfolio transitions executed in the U.S. equities market. Portfolio transition orders, which are proxies for bets, scale with an estimated exponent of -0.63 when measured as a fraction of trading volume; the predicted value is $-2/3$. Transaction costs have an estimated exponent of -0.32 ; the predicted value is $-1/3$. Kyle, Obizhaeva and Tuzun (2017), Bae et al. (2016), Andersen et al. (2015), and Benzaquen, Donier and Bouchaud (2016) generalize the invariance predictions from bets to trades and find cross-sectional and time-series invariance-implied relationships in the markets of U.S. equities, South Korean equities, the E-mini S&P futures, and other futures contracts. Kyle et al. (2010) document that the arrival of Thompson-Reuters news articles scale with an exponent close to $2/3$. Kyle and Obizhaeva (2017*a*) find that bid-ask spreads and number of transactions in the markets of U.S. equities and Russian equities are also consistent with quantitative predictions about how these variables relate to the measure of liquidity L .

While predictions of the invariance hypotheses do not fit the data exactly, they do fit the data

closely enough to suggest that the invariance hypotheses provide a natural yardstick for making comparisons in market microstructure quantities across many markets. Eugene Fama once said, “data still needs a theory it likes.” Our research implies that data seems to like most theoretical models of market microstructure when augmented by invariance assumptions. Without invariance, theoretical models generate predictions which are empirically difficult to test. The deviations from invariance provide a way to measure the effect of different institutional structures in different markets.

1 From 20th Century Block Trading to 21st Century Smooth Trading

The model of block trading assumes that large traders exercise market power by announcing a quantity, trading once, and then not trading again. In fact, after trading once, an informed trader with market impact has an incentive to trade again because the size of the first trade was restricted to limit price impact. Block trading is consistent with the Treynor model, the one period model of Kyle (1985), and the dynamic generalization described below. From the perspective of a trader, the price is a linear function of the trader's inventory; in this sense, holding an inventory S has permanent price impact, changing the price from p_0 with no inventory to $p_0 + \lambda S$, where λ measures permanent price impact.

The model of flow trading is consistent with the smooth trading model of Kyle, Obizhaeva and Wang (2018). In this model, traders participate in the market continuously, trading gradually so that inventories are a differentiable function of time. The price is a linear function of the trader's inventory and the time derivative of the trader's inventory. Compared with a price of p_0 if the trader holds no inventory and does not trade, holding inventory S and trading at rate $x = dS/dt$ drives the price to $p_0 + \lambda S + \kappa x$. There is temporary price impact because the price changes to $p_0 + \lambda S$ when a trader stops trading; the parameter κ measures temporary price impact.

This section reviews institutional and theoretical modeling principles that connect block trading models with smooth trading models.

Dealer Markets and Organized Exchanges. The defining feature of a dealer market is that customers are not allowed to place limit orders that can be matched with other customers limit orders in a manner that respects time and price priority. Instead of using the dealer as an agent (broker) to handle execution of an order, customers must trade with a dealer acting as principal. Such trading is not anonymous. Historically, trades have been negotiated by voice over the

telephone. Dealers may offer different bids and offers to different customers; they may also finalize prices after deals are negotiated. Manual processing of trades makes it difficult to enforce pre-trade and post-trade transparency.

The defining feature of an organized exchange is that customers can place limit orders that can be matched with other customers limit orders in a manner respecting time and price priority. Such matching can easily occur electronically. This makes it straightforward for the matching to be anonymous. Since computer interfaces have low marginal costs for handling additional messages, each transaction has a negligible order processing cost. Furthermore, the automated nature of the matching engine makes it easy to post bids and offers publicly and to publish traded quantities and price promptly; both pre-trade and post-trade transparency are straightforward.

Organized exchanges are economically more efficient than dealer markets because they realize network externalities and allow inexpensive interfaces to an electronic matching engine. The matching engine in an organized exchange allows small customers to access all bids and offers in the market at low cost. The high fixed costs of bargaining over the phone make it expensive for small customers to negotiate with multiple dealers. Therefore, small customers generally receive worse terms than large customers in a dealer market (see Harris (2015)).

In a dealer market, customers have an incentive to ask dealers to change their role to agents, giving the customers the same fast access to markets that dealers have. Competitive dealers have an incentive to respond to customer preferences by providing customers with the same pre-trade and post-trade transparency the dealers themselves have. To prevent these forces of competition from driving the market structure toward an organized exchange, dealer markets often have a hybrid structure, with an automated, transparent “inside” market for dealers, a nontransparent interface for customers, and cartel-like rules designed to prevent dealers from sharing their superior interfaces with customers.

Dealers as a whole benefit more from a dealer market than an organized exchange because the lack of transparency allows dealers to have better information about the state of the market than customers and high fixed costs for negotiating trades makes it costly for customers to negotiate better terms by playing one dealer off against another. In the absence of government regulation, market forces tend to drive the market structure toward a dealer market. Individual customers want their own traded quantities to be kept secret while observing other customers’ quantities traded and prices. When dealers accommodate this preference of customers, the market tends to become opaque.

Actual markets often have a hybrid structure combining features of an organized exchange with features of a dealer market. In a dealer market, insider markets for members of the club of dealers often resemble an organized exchange, with interdealer brokers insuring transparency

and dealer screens providing transparency to dealers. When organized exchanges are controlled by brokers and market makers, the rules of the exchange are often tweaked so that time and price priority are not fully respected and insiders, such as the NYSE specialist, have trading advantages over customers not present on the exchange floor.

In a dealer market conducted by telephone, dealers know the identity of each customer they trade with and each transaction has a high fixed cost because it takes time to negotiate terms of deals. Since dealers do not want informed customers going behind their backs by trading with competing dealers simultaneously, it is natural for dealer market transactions to be structured as block trades in which a customer trades a significant quantity with one dealer at one time and commits not to trade again for a significant amount of time. Even though the NYSE of the 1980s was officially an organized exchange on which traders could trade anonymously, many large trades were in fact negotiated “upstairs” as blocks in what effectively functioned as a dealer market, then crossed on the exchange floor. These dealer market practices are consistent with the block trading model in this paper, according to which informed traders trade only one time.

By contrast, since trading on an organized exchange is anonymous, large traders can execute large bets by placing many small orders over time. As electronic limit order books of the 21st century have replaced dealer markets of the 20th century, such “order shredding” strategies have displaced block trading. In the rest of this section, we first summarize the history of equity trading in the United States, then describe how theories of market microstructure can model the shift from block trading to order shredding.

History of U.S. Equity Trading. The history of equity trading in the United States illustrates economic differences between dealer markets and organized exchanges.

In the late 1700s, brokers in New York city entered into the Buttonwood agreement, which established a cartel designed to enforce fixed commissions. By organizing trading on a stock exchange with one physical location, the New York Stock Exchange not only achieved network externalities but also created a monitoring mechanism for preventing commission undercutting. In the 1970s, fixed commissions were declared illegal, and Nasdaq began to operate as an over-the-counter market with trading by telephone. In the 1980s and 1990s, the market design for stocks evolved as rivalry between dealer market structures like Nasdaq and organized exchanges like the NYSE. Trading on the NYSE took on characteristics of a dealer market, with large block trades negotiated “upstairs” (away from the exchange floor) like in a dealer market.

During the 1990s and early 2000s, changes in technology and regulation drove market structures away from dealer markets and toward organized exchanges. In the 1980s, both dealer markets and organized exchanges resisted the use of electronic technologies to match trades,

although the technology was implemented in other markets.¹ Continuing improvements in electronic technology made algorithmic trading possible. Since efficient algorithmic trading requires an electronic interface, this put pressure on stock markets to become more like an electronic exchange.

Research by Christie and Schultz (1994) pointed out the NASDAQ dealers seemed to avoid quoting prices in “odd-eighths,” a practice which raised the bid-ask spread on some stocks from 1/8 to 1/4 and thereby increased dealer profits at the expense of customers. Government regulators responded by forcing reforms in the market which introduced new order handling rules. The new rules required the dealers to accept customer limit orders and post them on a trading venue where other customer limit orders could trade against them. These new rules had the effect of converting NASDAQ from a dealer market to an organized exchange.

The tick size of one eighth of a dollar was cut to one sixteenth of a dollar and subsequently reduced to one cent in 2001. In the early 2000s, Regulation NMS encouraged competition among exchanges. This led to fragmented markets, with trading occurring on numerous exchanges, greatly undermined the monopoly profits exchanges had previously earned. After the new order handling rules were introduced, electronic trading rapidly replaced voice trading. Computer algorithms replaced high human market makers. Exchanges allowed traders pay fees for faster access to the market. Like the human market makers they replaced, the market-making algorithms acquired preferential access and exploited their speed to be profitable; they are frequently referred to as high frequency traders. Algorithmic trading and small tick size also made it economical for customers to break large block trades into many small pieces and execute the small pieces anonymously over time. Electronic order shredding changes the assumptions theoretical models need to make to model price formation.

Modeling of modern organized exchanges requires assumptions which allow customers to manage trading costs by placing many trades based on the same information. Since dynamic models can be mathematically complicated, it is worth examining the intuition that can be carried over from one-period models into a dynamic setting.

Noise Trading and Competition in One-Period Models. In the 1970s, economists’ understanding of speculative financial markets was governed by the ideology that markets for stocks, bonds, and commodity futures contracts represented close approximations to perfect competition.

¹It took a full generation for the human stock market of the 1980s to evolve into the electronic markets of the 21st century, even though electronic trading was occurring in the 1970s. West Texas farmers’ cooperative PCCA implemented electronic trading of spot of cotton with the TELCOT system in 1975 and electronic delivery of warehouse receipts in 1989. Electronic trading was implemented by the customers (a farmers’ cooperative) for the purpose of reducing the middle mens’ spread between the price farmers received and the price consumers (cotton spinning mills) ultimately paid. See Lindsey et al. (1991) and Lindsey et al. (1994); co-inventor Albert S. Kyle, III, is the father of this paper’s co-author.

Consistent with this tradition, Grossman and Miller (1988) present a model in which random noise demand immediacy by trading exogenous quantities with competitive market makers. A demand for immediacy goes hand in hand with assuming perfect competition because a price taker has, by definition, no incentive to constrain the quantity traded to reduce price impact. A price taker trades to his target inventory immediately.² This is consistent with a block trading model, in which a trader trades once to a target inventory and does not trade again until new information motivates a new trade. The size of the quantities traded are functions of the traders' risk aversion. Assuming exogenous noise traders trade a fixed quantity is a traditional short-cut consistent with the assumption of perfect competition.

The model of Kyle (1985) broke with the tradition of assuming financial markets are perfectly competitive by showing how a large trader with private information exercises market power by restricting the quantity traded to reduce impact. Since the informed trader is risk neutral, risk aversion does not constrain the quantity traded. Absence of risk aversion makes this a market microstructure model and not an asset pricing model. Nevertheless, exercise of market power by the large informed trader is consistent with Fama's concept of efficient markets because competitive, risk neutral market makers are assumed to extract information from the order flow and impound this information into prices. The model has exogenous noise traders, who trade impatiently, like perfect competitors, for some unmodeled exogenous motive.

The model assumes that the large informed trader's strategy is implemented as a quantity ("market order"), not a price contingent schedule ("limit order"). This assumption makes it easier to solve models analytically. The market order assumption is one way of capturing in a simple theoretical model the idea that markets are designed so that market makers have different and more flexible access to the market than traders who are not market makers.

An alternative approach is to assume that trading is a Bayesian game in which traders submit demand schedules that are functions of price. Kyle (1989) describes a one-period version of such a model in detail. The model assumes risk averse traders have exponential utility, some traders have private information and some do not, and noise traders trade exogenous random amounts. An auctioneer (nowadays a computer) calculates a market clearing price consistent with the orders submitted. A trader exercises market power by taking into account how the

²A demand for immediacy is suggested by the frenetic pace of trading on organized exchanges in the 1980s. While Grossman and Miller suggest that risk aversion motivates customers to shift risks quickly, a better explanation for the fast pace of activity on exchange floors derives from customer efforts to deal with moral hazard problems. Customers feared that unscrupulous brokers would use time delays in reporting trade executions or canceling orders to change favorable prices to unfavorable prices on executed orders or to allocate executions at favorable prices to other traders' accounts. Customers used public data feeds to monitor the quality of order executions by insisting the broker execute orders quickly and report trade executions and order cancellations quickly. Therefore, it is not surprising that the messengers who carried orders around the trading floor were called runners, even though rules technically prevented them from running.

demand schedule submitted affects the equilibrium price. While this modeling approach is analytically less tractable than assuming a market-order model, the solution to the model can be characterized in almost-closed form. The solution shows that assuming perfect competition makes quantities larger and prices more accurate because the informed traders trade more aggressively.

Lee and Kyle (2018) investigate strategic trading in a symmetric model which otherwise resembles the competitive model of Diamond and Verrecchia (1981). Instead of assuming exogenous random noise traders, it is assumed that informed traders have a hedging motive based on random endowment shocks. Since there are no uninformed trader, there is only one type of trader who has both private information and a hedging motive. This makes the model an analytically convenient device for studying the effect of strategic trading with imperfect competition. Since both hedging and trading on private information are affected symmetrically by exercise of market power, prices are not affected by the number of informed traders or how the private information is distributed among traders; imperfect competition shows up in quantities, not prices. Competitive prices are the same as prices in a model of perfect competition. Markets become less competitive, market liquidity falls, and quantities are restricted more when endowment shocks become smaller relative to risk bearing capacity and when there is more private information in the market. If there is too much private information in the market, liquidity dries up completely and there is no trade. The market becomes competitive only in a limit in which an infinite number of traders compete by trading on the same information. Lee and Kyle (2017) show that when exogenous noise trading goes to zero in a market which would otherwise have no liquidity, the limit is infinitely non-competitive and the price is not fully revealing. This result is exactly the opposite of the conventional wisdom of Grossman and Stiglitz (1980) that perfect competition makes prices fully revealing when noise trading vanishes.

Dynamic Models. The one-period and continuous-time models of Kyle (1985) have assumptions approximating 20th century trading. The models make a sharp distinction between market makers and other traders. Customers are required to trade with market makers. Market makers are assumed to condition the quantities they trade on price while other traders are required to trade quantities which do not respond to price. This modeling assumption captures the idea that market makers have the ability to adjust more quickly than other traders. While the model assumes for simplicity that market makers break even, it can easily be modified to allow market makers earn profits by trading at prices unfavorable to customers; this would show up as negative first-order autocorrelation in returns. Exogenous noise traders trade random, uncorrelated quantities over time. They do not attempt to reduce market impact costs by trading

strategically.³ This assumption is realistic if noise traders are small, almost-perfect competitors who face high fixed costs for placing an order.

The dynamic model of Kyle (1985), the large informed trader does not trade once in a large block but instead spreads his trading out over time; information is gradually incorporated into prices. Vayanos (1999) discusses a discrete-time model of gradual trading as a result of inventory shocks. This practice is consistent with an organized exchange, like the S&P E-mini future market, in which traders can submit orders anonymously to the exchange floor. While the forced intermediation of the market-order model is like a dealer market, the anonymity assumption is like an organized exchange.

In fact, many customers trade large quantities and have incentives to reduce market impact costs. In the absence of fixed costs for placing orders, large noise traders have an incentive to break large trades into small pieces, trading at improved prices by gradually walking along a market supply schedule rather than trading an entire block at a price which reflects maximum impact. If all noise traders act strategically in this manner, the equilibrium with immediate block trading breaks down. There is no instantaneous depth; market depth is only made available over time.

The smooth trading model of Kyle, Obizhaeva and Wang (2018) is a well-specified model of dynamic equilibrium trading by optimizing, risk-averse informed traders on a continuous organized exchange. Each trader has a continuous flow of new information, is present in the market at all time, and optimizes trading on private information over time to manage market impact costs optimally. Instead of being motivated by noise trading, trading is motivated by relative overconfidence: each trader believes his flow of private information is more precise than the flow of private information to other traders.⁴ In all other respects, the equilibrium is a Bayesian Nash equilibrium.

In one-period models, imperfectly competitive traders trade a fraction of the quantity that a competitive trader would trade. In this model of smooth trading, each trader exercises market power by trading slowly; equilibrium inventories adjust continuously toward target inventories which constantly change as a result of new information arriving continuously. There is no instantaneous market liquidity; trading a given quantity takes time. A single-price auction clears the market continuously for flow-quantities defined as the time derivatives of traders' inventories. In the competitive model of Kyle, Obizhaeva and Wang (2017), which otherwise has similar structure, there is instantaneous market liquidity and traders trade immediately to target inventories. Because of the assumption of overconfidence, trading has the flavor of a Keyne-

³See Black (1995)

⁴Kyle and Wang (1997) show that overconfidence can be a commitment device to improve speculative profits by making other traders trade less aggressively in a duopoly model.

sian beauty contest; the beauty contest is more pronounced with markets are more liquid and is most pronounced in a competitive setting.

The setting of the smooth trading equilibrium approximates an organized exchange in which traders trade continuously by submitting flow demand schedules. In today's markets, order shredding algorithms approximate smooth trading by breaking orders into tiny pieces and submitting them over time. As a result of market fragmentation, a small minimum tick size of one cent, and various other market rules, these order-shredding algorithms place and cancel many orders for small quantities and thus incur high bandwidth costs. Since high frequency traders have faster technology and lower bandwidth costs than other traders, they have been able to earn profits for themselves by exploiting their enormous bandwidth and superior speed.⁵

Budish, Cramton and Shim (2015) argue that inefficiencies associated with this technological arms race can be dramatically reduced by using frequent batch auctions to make trading more discrete. Consistent with the theory of smooth trading, Kyle and Lee (2017) propose instead that exchanges implement a new order type called continuous scaled limit orders, which directly enable flow-trading by making orders execute gradually over time at a maximum rate specified by the order. These orders make organized exchanges function more efficiently for two reasons. First, they reduce the message costs incurred by order-shredding algorithms by allowing one simple order to replace many thousands of small orders and cancelations. Second, they virtually eliminate the speed advantages high frequency traders have over slower traders by dramatically reducing the quantities high frequency traders can trade by being slightly faster than other traders. Implementing these changes will make stock market trading resemble closely what Black (1971) imagined.

While these models describe theories connecting informed speculation with market liquidity and the information content of prices, the models are difficult to test empirically because the models themselves do not make clear how to map the theoretical concepts like permanent price impact λ and temporary price impact κ into data on prices, trading volume, and returns volatility. The remainder of this paper describes a meta-model which can be used to derive an empirically-oriented measure of liquidity and map this measure into model parameter like λ and κ , thus connecting market liquidity to adverse selection in a simple and straightforward manner.

⁵In a study of market making during the flash crash of 2010, Kirilenko et al. (2017) document that a small group of high-frequency traders accounted for about 30 percent of trading volume. They also traded smaller quantities, suggesting lower bandwidth costs.

2 Bets and Microstructure Variables

There are many market microstructure variables. Our modeling approach pays rigorous attention to consistency of units. This section introduces notation for describing market microstructure variables, illustrated with typical values and units of measurement.

Easy-to-observe quantities include price, volume, and volatility:

$$\begin{aligned} \text{Price} &= P = 40.00 \text{ dollars/share}, & (1) \\ \text{Trading Volume} &= V = 1.00 \text{ million shares/day}, \\ \text{Returns Variance} &= \sigma^2 = 0.02^2/\text{day}, \\ \text{Volatility} &= \text{Returns Standard Deviation} = \sigma = 0.02/\text{day}^{1/2}. \end{aligned}$$

To understand what follows, it is crucial to keep in mind that returns “volatility” has units $1/\sqrt{\text{day}}$.

The term “bet” describes approximately uncorrelated decisions to purchase or sell specific institutional quantities of financial assets. In the 20th century, it was common to implement a bet as one block trade. In the 21st century, bets typically correspond to meta-orders, which are implemented with child orders generating many small trades executed algorithmically over time. The rate at which bets arrive in a market measures the speed with which business time passes in that market. Variables which are hard to measure and vary greatly across assets include bet size, number of bets, and the price impact coefficient:

$$\begin{aligned} \text{Size of Bet} &= Q = 10\,000 \text{ shares} & (2) \\ \text{Number of Bets} &= \gamma = 100/\text{day} \\ \text{Execution Horizon} &= H = 1 \text{ day} \\ \text{Price Change per Bet} &= \Delta P = 0.04 = \text{dollars/share} \\ \text{Percentage Price Change per Bet} &= G = 10 \text{ basis points} = 10^{-4} \text{ (dimensionless)} \\ \text{Price Impact Coefficient} &= \lambda = 5 \times 10^{-5} \text{ dollars/share}^2 \\ \text{Fundamental Value} &= F = 150 \text{ dollars/share} \\ \text{Price Error} &= \Sigma^{1/2} = \text{Var}^{1/2} \left[\ln \left(\frac{F}{P} \right) \right] = \ln(2) \text{ (dimensionless)} \\ \text{Price Resiliency} &= \rho = 0.0040/\text{day} \text{ (dimensionless)} \end{aligned}$$

Variables which are hard to measure but which we hypothesize do not vary much across assets

include the following:

$$\text{Average Dollar Impact Cost per Bet} = C = 2000 \text{ dollars}, \quad (3)$$

$$\text{Moment ratio} = m := \frac{E[|Q|] \sqrt{E[|Q|^{2\beta}]}}{E[|Q|^{\beta+1}]} = \sqrt{2/\pi} \text{ (dimensionless)}$$

$$\text{Moment ratio} = m_\beta := \frac{(E[|Q|])^{\beta+1}}{E[|Q|^{\beta+1}]} = \sqrt{2/\pi} \text{ (dimensionless)}$$

$$\text{Probability a bet is informed} = \theta = 1/2 \text{ (dimensionless),}$$

$$\text{Information content of a private signal} = \tau = 0.0080^2 \text{ (dimensionless signal-to-noise ratio).}$$

$$\text{Information content of a bet} = \theta^2 \tau = 0.0040^2 \text{ (dimensionless signal-to-noise ratio),}$$

$$\text{Cost of an Informative Signal} = \bar{c}_I = 2000 \text{ dollars.}$$

The information content of a private signal is greater than the information content of a bet because some bets are based on noise—not private signals with information content—and the market cannot tell the difference between bets based on information and bets based on noise.

Other variables potentially of interest include the following:

$$\text{Bid-Ask Spread} = S = 0.05 \text{ dollars/share}, \quad (4)$$

$$\text{Minimum Tick Size} = K_{\text{MIN}} = 0.01 \text{ dollars/share},$$

$$\text{Minimum Lot Size} = Q_{\text{MIN}} = 100 \text{ shares},$$

$$\text{Trade Size} = X = 200 \text{ shares},$$

$$\text{Number of Trades} = N = 50\,000/\text{day}.$$

The variables differ in terms of how easy it is to obtain their values. Some variables such as trading volume V and prices P are routinely reported by exchanges. Other variables such as returns volatility σ are straightforward to estimate from publicly available data. Information about tick sizes and lot sizes can be looked up on the web sites of exchanges.

Many parameters can be only inferred indirectly or obtained with great difficulty from proprietary data sets. For example, even though it is usually possible to observe executed trades, information on intended meta-orders and features of execution algorithms is not easily available to researchers. The estimation of transaction costs is also complicated by issues of endogeneity and statistical power. Some variables—such as pricing accuracy, precision of private information, and fraction of informed traders—are difficult to infer even from proprietary data.

3 Meta-Model and Invariance

Consider the following simple meta-model, which describes how bets are related to trading volume and returns volatility. Suppose bets of random size Q arrive randomly into the market at rate γ (per day). The bets are approximately independently distributed, with positive values of Q denoting buying and negative values denoting selling. While each bet represents a new investment idea based on private information, the market only sees trades which implement the bets, not the information upon which each bet is based. A bet of size Q moves the price by ΔP (dollars per share). We assume this price impact to be a power function of the bet's size Q :

$$\Delta P := \lambda Q^\beta. \tag{5}$$

The dimensionless exponent β describes the shape of the market impact function. Linear impact, consistent with many theoretical models, implies $\beta = 1$. A square root impact function, consistent with empirical research, implies $\beta = 1/2$. The parameter λ , which defines “market impact,” has units dollars/share^{1+ β} .

The following three-equation “meta-model” describes how bets and their price impact affect share volume V (shares per day) and returns volatility σ (per 1/day^{1/2}):

$$\begin{aligned} V &= \gamma E[|Q|] && \text{(Definition of Volume),} && (6) \\ E[(\Delta P)^2] &= \lambda^2 E[|Q|^{2\beta}] && \text{(Market Impact of One Bet),} \\ \sigma^2 &= \gamma E\left[\left(\frac{\Delta P}{P}\right)^2\right] && \text{(Bets Generate All Volatility).} \end{aligned}$$

The first equation, “definition of volume,” says that trading volume is the sum of all bets. It implicitly assumes that the other side of each bet is taken by an intermediary (market maker) which is not placing a bet. The second equation “market impact of one bet” says that the variance of the market impact ΔP of one bet is consistent with equation (5). The third equation “bets generate all volatility” results from assuming that bets are independently distributed and have permanent price impact ΔP from equation (5).

These equations can be easily modified to accommodate more intermediation volume and public announcements which move prices. The volume V on the left side of the first equation can be reduced by the fraction of trading that represents intermediaries trading with intermediaries and increased when a buy bet trades against a sell bet without intermediation. The variance on the left side of the third equation can be reduced by the percentage of returns variance that results from public announcements that move prices without trading.

The system of equations (6) is called a meta-model because many models of market mi-

microstructure are consistent with the three equations, while a complete model of market microstructure is likely to have a richer structure than captured by these three equations alone. In the system (6), the three quantities measuring volume, volatility, and price are easy to measure:

$$V, \quad \sigma, \quad P \quad (\text{Easy to Measure}). \quad (7)$$

The five remaining quantities—measuring bet-arrival rate, two moments of bet size, and two measures of price impact—are hard-to-measure:

$$\gamma, \quad E[|Q|], \quad E[|Q|^{2\beta}], \quad E[(\Delta P)^2], \quad \lambda \quad (\text{Hard to Measure}). \quad (8)$$

These five quantities are hard to measure because it is difficult to infer statistically the size of bets, the number of bets, and the price impact of bets from transaction data. Especially in the 2000s, large bets are often executed as sequences of small trades, which makes it difficult for researchers to reconstruct the size of bets from transactions data. The hard-to-measure quantities are likely to vary greatly across assets. For example, stocks with relatively higher levels of trading volume are likely to have more bets and larger bets executed with lower market impact in a given day.

Since the three-equation meta-model (6) imposes three log-linear restrictions on the five hard-to-measure unknowns, two more restrictions are needed to solve for the hard-to-measure unknowns. Without two more restrictions, the three-equation meta-model (6) does not make it possible to infer values of the five unknown quantities.

To solve the problem of having too few equations in too many unknowns, we adopt the following strategy, which we call “market microstructure invariance.” We add to the system a new equation with a new unknown. We choose the new unknown to be a quantity which is unlikely to vary much, if at all, across assets and time. If the quantity is sufficiently invariant, then its value needs to be calibrated only once from data. The data used for calibration may be special data which makes it possible to identify individual bets. The calibrated value can then be added to the system as a known constant, reducing the difference between the number of equations and unknowns by one.⁶

Let C denote the ex ante unconditional expected dollar cost $E[|Q \Delta P|]$ of executing a bet:

$$C = E[|Q \Delta P|] = \lambda E[|Q|^{1+\beta}]. \quad (9)$$

Transaction costs invariance hypothesizes that C is invariant across assets and time. Taken less strictly, the hypothesis says that C may vary somewhat across assets, but not nearly as much as

⁶This approach is used by Kyle and Obizhaeva (2016a) and later clarified in Kyle and Obizhaeva (2018).

the number of bets γ or the average bet size $E[|Q|]$.

Indeed, Kyle and Obizhaeva (2016a) use data on portfolio transitions to show that the average dollar cost of executing bets is approximately constant across assets, with value $C \approx \$2,000$. The invariance conjecture then implies that the average cost of a bet is approximately \$2,000 for all markets—equities, fixed income, foreign exchange, or commodities.

Transaction cost invariance is motivated by the following economic intuition. To beat a benchmark, an asset manager must make expected trading profits which cover both the real resource costs of acquiring costly information and the market impact costs of trading on it. Finance professionals allocate their scarce skills and resources across different markets to maximize the value of trading on the costly signals generated by their research. In the search for markets with better profit opportunities, traders shift their attention from one market to another. If generating trading ideas and implementing them turn out to be cheaper in some markets, then more traders flock onto these opportunities. In contrast, if generating and executing bets are more expensive, then traders are likely to leave the market in search for better investment opportunities elsewhere. In the resulting dynamic equilibrium, the salaries of traders, ex ante costs of generating ideas, costs of executing bets, and expected profits on each investment become equal across markets. This equality of equilibrium dollar costs per bet across markets may justify our assumption about invariance of C . It is also consistent with the hypothesis that the cost of an informative signal, denoted \bar{c}_B , is proportional to, and perhaps the same as, the market impact cost of a bet C .

Adding the additional equation leads to a system of four equations:

$$\begin{aligned}
 V &= \gamma E[|Q|] && \text{(Definition of volume),} && (10) \\
 \sigma^2 &= \gamma E\left[\left(\frac{\Delta P}{P}\right)^2\right] && \text{(Bets generate all volatility),} \\
 E[(\Delta P)^2] &= \lambda^2 E[|Q|^{2\beta}] && \text{(Market impact of one bet),} \\
 C &= \lambda E[|Q|^{1+\beta}] && \text{(Dollar impact cost of a bet).}
 \end{aligned}$$

Assuming the invariant quantity C and the easy-to-measure quantities are known, the known quantities are P , V , σ , and C . The fourth equation in (10) contains the new unknown value $E[|Q\Delta P|]$. Adding this new value to the five in (8) results in six hard-to-measure quantities. Therefore, we have four equations in six unknowns. We still need to add two restrictions to be able to solve the system.

To address this problem, notice that three of the unknown quantities are moments of the distributions of bet sizes Q : $E[|Q|]$, $E[|Q|^{2\beta}]$, and $E[|Q|^{1+\beta}]$. The quantity $E[(\Delta P)^2]$ has moments of Q from the definition $\Delta P = \lambda|Q|^\beta$ in (5).

The two needed log-linear restrictions are obtained by assuming that the distribution of Q is the same across assets except for a scaling factor that defines the mean $E[|Q|]$. We call this restriction *bet size invariance*, consistent with Kyle and Obizhaeva (2016a). This conjecture makes the relevant moments of Q invariant. Instead of making this strong invariance assumption, we could have made the weaker assumption that two specific moment ratios of Q are invariant. The two specific invariant moment ratios, denoted m and m_β are defined by

$$m := \frac{E[|Q|] \sqrt{E[|Q|^{2\beta}]}}{E[|Q|^{\beta+1}]}, \quad m_\beta := \frac{(E[|Q|])^{\beta+1}}{E[|Q|^{\beta+1}]}.$$
 (11)

Since the exponents of $|Q|$ in the numerator and denominator are the same in both definitions, the moment ratios m and m_β will be constant across assets as long as any two distributions of bet sizes Q differ across assets and time only by a scalar.

Using portfolio transition orders in U.S. equities as proxies for bets, Kyle and Obizhaeva (2016a) calibrate estimates of the distribution of bet size by pooling data across many stocks. The distribution of unsigned portfolio transition order size closely resembles a log-normal distribution with a log-variance of approximately 2.50. Signed orders are approximately symmetric about zero (in the sense that buy-bets and sell-bets of a given size are equally likely). The moment ratios m and m_β are functions of the log-variance of the lognormal distribution: $m = \exp(2.50(\beta^2 - 2\beta)/2)$ and $m_\beta = \exp(-2.50\beta(\beta + 1)/2)$. If $\beta = 1/2$, then $m_\beta = m \approx 0.40$. If $\beta = 1$, then $m \approx 0.25$ and $m_\beta = m^2$.

Equations (10)–(11) are a system of six log-linear equations in terms of six hard-to-measure unknown quantities:

$$\gamma, \quad E[|Q|], \quad E[|Q|^{2\beta}], \quad E[|Q|^{1+\beta}], \quad E[\Delta P^2], \quad \lambda.$$
 (12)

These equations can be solved for the six unknowns in terms of the three easily observable characteristics P , V , and σ , which vary greatly across assets, and the three hard-to-observe characteristics C , m , and m_β , which are likely to be approximately constant across assets. In terms of these characteristics, the solutions for $E[|Q|]$, γ , and λ are

$$E[|Q|] = \left(\frac{mC}{\sigma PV} \right)^{2/3} V$$
 (13)

$$\gamma = \left(\frac{\sigma PV}{mC} \right)^{2/3},$$
 (14)

$$\lambda = m_\beta \left(\frac{C\sigma^2}{m^2 PV} \right)^{(1+\beta)/3} \frac{P^{\beta+1}}{C^\beta}.$$
 (15)

Using equations (11)) for moments of bet size and the definition of ΔP in (5), one can easily obtain the solution for the other three unknowns $E[|Q|^{2\beta}]$, $E[|Q|^{1+\beta}]$, and $E[(\Delta P)^2]$ as well. These predictions are empirically operational. The proportionality constants determined by C , m , and m_β can be calibrated from pooled data on special data sets. These calibrated values can be combined with values for P , V and σ from readily available data sets like the CRSP and TAQ databases for the U.S. equities market. Values for hard-to-measure quantities can be calculated as solutions to the six equations in six unknowns (13)–(15).

The meta-model restrictions, augmented with invariance hypotheses, imply that order flow has a very particular structure and allow it to be described using observable dollar volume and volatility. The exponents of 1/3 and 2/3 in equations (13) and (14) show how the size and number of bets in the order flow vary with volume and volatility given the transaction costs invariance assumption. Holding price and volatility constant, a larger trading volume PV is 1/3 due to larger average bet size $E[|PQ|]$ and 2/3 due to a larger number of bets γ . For example, an increase in dollar volume PV by a factor of 8 is predicted to lead to an increase in average dollar bet size by a factor of 2 ($= 8^{1/3}$) and to an increase in the number of bets by a factor of 4 ($= 8^{2/3}$). The solution also describes how market depth λ depends on dollar volume and volatility.

To summarize, specifying generic features of theoretical models, without filling in specific details regarding private information or adverse selection, makes it possible to obtain operational predictions about order flow and liquidity.

Business Time. It is also possible to obtain the same results by augmenting the system (6) with an alternative invariance conjecture regarding the size of bets rather than their market impact.⁷ If the expected arrival rate of bets γ sets the pace of business time in which markets function, then the dollar returns standard deviation per unit of business time is $P\sigma/\sqrt{\gamma}$. *Bet size invariance* can alternatively be defined by hypothesizing that the dollar risks that bets Q transfer per unit of business time,

$$I := Q \frac{P\sigma}{\sqrt{\gamma}}, \quad (16)$$

have an invariant mean $E[|I|]$ for all markets. If one assumes that not only the mean $E[|I|]$ but also the entire distribution of I is invariant across markets, the invariance of the moment ratios m and m_β is implied as well. We could refer to this hypothesis as the “strong invariance” hypothesis.

It can be shown that the transaction cost and bet size invariance hypotheses are closely related to each other, since the dollar volatility per bet $P\sigma/\gamma^{1/2}$ is closely related to expected price change per bet ΔP . Indeed, the definition of m and power market impact function $\Delta P =$

⁷Kyle and Obizhaeva (2016a) introduce this approach and describe it in more detail.

$\lambda|Q|^\beta$ yield

$$E[|I|] = E\left[|Q| \frac{P\sigma}{\sqrt{\gamma}}\right] = E\left[|Q| \left(E[(\Delta P)^2]\right)^{1/2}\right] = m C. \quad (17)$$

The average dollar cost $C = E[|Q| \Delta P]$ is equal to the product of average bet size $E[|Q|]$ and average dollar impact $P\sigma/\sqrt{\gamma}$ per bet, with the factor m adjusting for differences in moments of the distribution of Q .

4 Dimensional Analysis, Leverage Neutrality, and Invariance

Another equivalent approach for obtaining similar operational predictions about market microstructure variables is based on applying dimensional analysis, leverage neutrality, and invariance. This approach makes it clear where the exponent of 1/3 comes from.

In finance, dimensional analysis begins with fundamental units of time, value, and asset quantity. Natural conservation laws take the form of the no-arbitrage restrictions, one of which is the principle of leverage neutrality.⁸

Start with the assumption that the variables of interest— $E[|Q|]$, γ , and λ —are determined by some unknown functions f_q , f_γ , and f_λ , which take share volume V , share price P , returns volatility σ , and expected dollar costs C as their arguments:

$$E[|Q|] = f_q(V, P, \sigma^2, C), \quad (18)$$

$$\gamma = f_\gamma(V, P, \sigma^2, C), \quad (19)$$

$$\lambda = f_\lambda(V, P, \sigma^2, C). \quad (20)$$

Second, reduce the dimensionality of the problem by forming dimensionless parameters from ratios of dimensional parameters. Share volume V is measured in shares-per-day, share price P is measured in dollars-per-share, returns variance σ^2 is measured per day, and average costs C is measured in dollars. The Buckingham π -theorem implies that functions f_q , f_γ , and f_λ can be represented by multiplying some product of powers of parameters that match the dimensions of the variable of interest⁹ by some unknown function of a smaller number of variables, equal to the number of parameters (four) minus the number of dimensions (three). In this case the single dimensionless variable can be taken to be $\frac{C\sigma^2}{VP}$. Here, to simplify exposition, we assume that the unknown functions are a power function with exponents α_q , α_γ , and

⁸The details of this approach are presented in Kyle and Obizhaeva (2017a).

⁹The dimensions of $E[|Q|]$ are the same as C/P , the dimensions of γ are the same as σ^2 , and the dimensions of λ are the same as $P^{\beta+1}/C^\beta$.

α_λ , respectively:

$$E[|Q|] = \frac{C}{P} g_q\left(\frac{C\sigma^2}{VP}\right) \sim \frac{C}{P} \left(\frac{C\sigma^2}{VP}\right)^{\alpha_q}, \quad (21)$$

$$\gamma = \sigma^2 g_\gamma\left(\frac{C\sigma^2}{VP}\right) \sim \sigma^2 \left(\frac{C\sigma^2}{VP}\right)^{\alpha_\gamma}, \quad (22)$$

$$\lambda = \frac{P^{\beta+1}}{C^\beta} g_\lambda\left(\frac{C\sigma^2}{VP}\right) \sim \frac{P^{\beta+1}}{C^\beta} \left(\frac{C\sigma^2}{VP}\right)^{\alpha_\lambda}. \quad (23)$$

The exponents α_q , α_γ , and α_λ are undetermined so far.

Third, assume that the functions satisfy a leverage neutrality restriction. Since a cash-equivalent asset embeds no risk, it can be exchanged in the market at no cost. This implies that the market impact cost of exchanging a risky bundle of assets is the same for any positive or negative amount of cash-equivalent assets included with the bundle. Changing leverage by a factor a does not change Q , γ , V , C , but changes P to Pa , σ^2 to $\sigma^2 a^{-2}$, and λ to λa . This restriction implies that $\alpha_q = -1/3$, $\alpha_\gamma = -2/3$, and $\alpha_\lambda = (\beta + 1)/3$:

$$E[|Q|] \sim \frac{C}{P} \left(\frac{C\sigma^2}{VP}\right)^{-1/3}, \quad (24)$$

$$\gamma \sim \sigma^2 \left(\frac{C\sigma^2}{VP}\right)^{-2/3}, \quad (25)$$

$$\lambda \sim \frac{P^{\beta+1}}{C^\beta} \left(\frac{C\sigma^2}{VP}\right)^{(\beta+1)/3}. \quad (26)$$

The equations above are identical to equations (13)–(15) implied by the solution to the meta-model, up to the missing proportionality factors depending on m and m_β .

In comparison to the dimensional analysis approach, the meta-model pins down some dimensionless proportionality constants. Dimensional analysis helps to define what we learn simply from mechanically matching dimensions and what we learn from deeper economic restrictions related to no-arbitrage principles. For example, the cube root is obviously related to no-arbitrage restrictions in the context of dimensional consistency.

This approach crucially depends on a correctly specified “guess” of the needed parameters in the first step. The invariance assumption is an important ingredient for empirical application because the cost C shows up in the lists of arguments (P, V, σ^2, C) for all variables of interest. Kyle and Obizhaeva (2018) emphasize that if C is omitted from the list of arguments, then dimensional analysis usually leads to results that are inconsistent with leverage neutrality and empirically unrealistic. The notable exception is the case of market impact function. If one simply assumes that $\Delta P(Q)/P = f(Q, P, V, \sigma^2)$, then dimensional analysis and leverage neutrality imply the square root market impact function $\Delta P(Q)/P = \sigma \sqrt{Q/V}$, which is often used by

practitioners¹⁰.

5 Solution Using Liquidity Measure L

The solution obtained above using the two general approaches can be conveniently described using the liquidity measure implied by invariance.

It can be shown that multiplying PV by m^2 in the definition of L scales the reciprocal $1/L$ so that $1/L$ is the ratio of the expected dollar cost of executing a bet C to the expected dollar volume of the bet $E[|PQ|]$:

$$\frac{1}{L} := \frac{C}{E[|PQ|]} = \left(\frac{\sigma^2 C}{m^2 PV} \right)^{1/3}. \quad (27)$$

The last equation is obtained using equation (13).¹¹

Intuitively, $1/L$ denotes the dollar-weighted average market impact cost of executing bets expressed as a fraction of the dollar value traded. For example, suppose that 1/3 of dollar volume consists of numerous small bets which cost 10 basis points each, and the other 2/3 of dollar volume consists of a few large bets which cost 40 basis points each. Then the dollar-weighted average cost of a bet in this market is 30 basis points $((1/3) \times 10 + (2/3) \times 40)$, so we have $1/L = 0.0030$.

Up to scaling, the illiquidity measure $1/L$ is the only dimensionless variable that can be constructed out of P, V, σ^2 , and C and has the same sensitivity to leverage as σ . It is invariant to changes in value units (from dollars to cents), quantity units (from shares to contracts), and time units (from days to second). It is also invariant with respect to a sampling frequency. These characteristics are rarely observed for other variables in finance. For example, the liquidity measure of Amihud (2002) has the same time units as $1/\sigma$.¹²

The solution for $E[|Q|]$, γ , and λ in (26) and (26)–(15) as well as (24)–(26) can be expressed in a simple manner using L as

$$E[|PQ|] = CL, \quad (28)$$

$$\gamma = \frac{1}{m^2} \sigma^2 L^2, \quad (29)$$

$$\lambda = \left(\frac{1}{L} \right)^{1+\beta} C^{-\beta} m_\beta P^{\beta+1} = m_\beta \frac{P}{L} \left(\frac{P}{CL} \right)^\beta. \quad (30)$$

¹⁰See Barra (1997) and Grinold and Kahn (1995)

¹¹From calibrations in Kyle and Obizhaeva (2016a), C is approximately equal to \$2000 and m is approximately equal to 0.25. A more liquid market is associated with higher dollar volume PV and lower volatility σ .

¹²The illiquidity measure $ILLIQ$ of Amihud (2002) is proportional to $\sigma (PV)^{-1}$; invariance implies that this measure is proportional to L^{-3}/σ .

Under the assumption m and C are invariant, these equations imply that when liquidity L goes up, the number of bets γ increases twice as fast as average bet size $E[|PQ|]$.

The solution for the market impact function $\Delta P = \lambda|Q|^\beta$ can be expressed in a simple manner using L as well.¹³ Let Z define bet size scaled by its mean CL , and let f be a specific invariant function of Z :

$$Z := \frac{Q}{E[|Q|]} = \frac{PQ}{CL} \quad \text{and} \quad f(Z) = m_\beta |Z|^\beta. \quad (31)$$

Then, we can then write the market impact formula $G = \Delta P/P$ where $\Delta P = \lambda|Q|^\beta$ as the product of two factors $1/L$ and $f(Z)$:

$$G = \frac{\Delta P}{P} = \frac{\lambda}{P} |Q|^\beta = \frac{1}{L} f(Z). \quad (32)$$

Equation (32) is a universal market impact function which can be applied to all market based on knowing only three invariant constants C , m , and m_β . It nests the square root model ($\beta = 1/2$) and linear price impact model ($\beta = 1$). It shows that percentage market impact G can be decomposed into the product of an asset-specific illiquidity index $1/L$ and an invariant function $f(\cdot)$ of scaled bet size $Z := Q/E[|Q|]$, which measures the size of a bet relative to an average bet $E[|PQ|] = CL$. The moment ratio m scales the definition of $1/L$ so that it measures average market impact cost, and the moment ratio m_β scales the function $f(Z)$. The percentage market impact G is substantial, if asset illiquidity is high and a bet is large relative to typical bets executed in that market.¹⁴

A Numerical Example. To illustrate, assume linear market impact $f(Z) = m_\beta Z$, with $m = 0.25$ and $m_\beta = m^2 = 0.0625$. Consider two markets. The first market has $1/L = 0.0100$ or 100 basis points; it is consistent with equation (27) with the assumptions $C = \$2000$, volatility $\sigma = 2\%$ per day, and dollar volume $PV = \$12,800,000$ per day. In this market, a bet of average size ($Z = 1$) has a market impact of 6.25 basis points, and a bet ten times larger than average ($Z = 10$) has a market impact of 62.5 basis points using $G = 0.000625Z$. The second market is twice less illiquid with $1/L = 0.050$ or 50 basis points; it can be obtained by multiplying dollar volume by a factor of 8 in the first example, so that dollar volume satisfies $PV = \$102,400,000$ per day. In this market, the average bet ($Z = 1$) has a market impact of 3.125 basis points, and a bet ten times larger than average ($Z = 10$) has a market impact of 31.25 basis points using $G = 0.0003125Z$.

What L Measures. During the financial crisis, volatility increased substantially, and dollar trading volume in some securities dropped. It is universally agreed that liquidity fell. Quan-

¹³The following exposition uses notation from Kyle and Obizhaeva (2017a)

¹⁴We have presented our derivation under the assumption of power market impact functions, but a similar decomposition (32) can be derived for any functional form g for asset-specific market impact functions. The more general case implies $f(Z) := g(Z)/E[|Z|g(Z)]$, with $g(Z) = |Z|^\beta$ for the power function.

tifying how much liquidity fell requires a precise understanding of what “liquidity” is exactly supposed to measure. Since L is a dimensionless number precisely calibrated for measuring market impact costs, it is straightforward to quantify different ways of measuring how much transactions costs fell:

- The dollar-weighted average market impact cost of executing bets of the size they are actually executed in the market, expressed as a fraction of the dollar value traded, is exactly equal to $1/L$.
- The percentage bid-ask spread may be proportional to $1/L$.
- The cost (in basis points) of exchanging some fixed dollar notional value (such as $PQ = \$1$ million) is proportional to L^{-2} for linear price impact and proportional to $L^{-3/2} \sim \sigma/(PV)^{1/2}$ for square root price impact.
- The cost (basis points) of transacting a fixed fraction of average daily volume (such as constant Q/V) is proportional to σ^2/L for linear price impact and σ for square root price impact.

In the following sections, we will explain that the measure of liquidity L is also closely related to resiliency, pricing accuracy, and different aspects of funding liquidity.

6 The Model of Treynor (1995)

Treynor (1995) describes a simple theoretical model of the bid-ask spread. In practice, the bid-ask spread reflects transaction costs of retail traders who execute small orders, which do not move prices beyond the best bid and ask prices. Treynor’s model is populated by informed traders, noise traders, and dealers who intermediate between asynchronous buyers and sellers in a presence of asymmetric information. Block trades of informed traders and noise traders move prices because of their potential information content; price impact results from adverse selection, and the dealer half-spread measures the market impact of block trades.

Let P denote the bid-ask spread midpoint, which measures the initial expected long-term fundamental value or liquidation value of the asset. Let S denote the dealer spread; the bid and ask prices are

$$B := P - \frac{1}{2}S, \quad \text{and} \quad A := P + \frac{1}{2}S. \quad (33)$$

Institutional investors buy or sell blocks of size $Q = \pm \bar{Q}$ shares; for simplicity, all bets are the same unsigned size. A dealer (specialist or market maker) sets the bid and ask prices to reflect private information contained in the bets placed by institutional investors. With probability

θ , a bet is placed by an informed investor who knows that the fundamental value of the asset differs from the current price P by ΔF per share. With probability $1 - \theta$, the bet is placed by a noise trader (liquidity trader), in which case the direction of the trade is uncorrelated with the difference between price and fundamental value. The dealer liquidates inventories at some distant future date, obtaining in expectation the fundamental value of the asset.

The dealer sets a bid price B for sell bets and ask price A for buy bets, so that on average they balance out losses on trading with informed traders against profits on trading with noise traders:

$$\frac{\theta}{2}((P + \Delta F) - A) = \frac{1 - \theta}{2}(A - P), \quad (34)$$

$$\frac{\theta}{2}(B - (P - \Delta F)) = \frac{1 - \theta}{2}(P - B). \quad (35)$$

These break-even conditions imply $B = P - \theta\Delta F$ and $A = P + \theta\Delta F$. Thus, the bid-ask spread is

$$A - B = 2\theta \Delta F. \quad (36)$$

The model implies that the endogenous percentage bid-ask spread is equal to the probability of a trader being informed times how much informed traders believe that the market price deviates from fundamentals. In practice, the theoretical variables ΔF and θ are difficult to observe. At first glance, this would appear to make the Treynor model a nice story of limited practical value.

In the context of the meta-model, the Treynor model is very practical. It can be shown that the meta-model system (10) and (11) is perfectly consistent with the Treynor model if the dealer spread is interpreted as the price impact of a bet. This makes it possible to generate operational predictions. Since noise traders and informed traders place bets of size $Q = \pm \bar{Q}$ shares with equal probabilities, this makes the moment ratios both equal to one ($m = m_\beta = 1$). Since each time price moves up or down by the same amount, $P \pm S/2$, the model is consistent with linear price impact. Suppose there are γ bets per day. Then, the meta-model system can be written as

$$V = \gamma |Q| \quad (\text{Definition of volume}) \quad (37)$$

$$\sigma^2 = \gamma \left(\frac{\Delta P}{P} \right)^2 \quad (\text{Bets generate all volatility}), \quad (38)$$

$$\Delta P = \frac{S}{2} \quad (\text{Market impact of one bet}), \quad (39)$$

$$C = |Q| \Delta P \quad (\text{Dollar impact cost of a bet.}) \quad (40)$$

$$(41)$$

This four-equation system has four unknowns γ , Q , ΔP , and S . The solution for the half percentage spread $S/(2P)$ is

$$\frac{1}{2} \frac{S}{P} = \left(\frac{\sigma^2 C}{PV} \right)^{1/3} = \frac{1}{L}. \quad (42)$$

A comparison with equation (27) (assuming $m = 1$) implies that the percentage half-spread is equal to the liquidity measure $1/L$.

For example, suppose an empirical researcher observes a price P of \$40.00 per share, volume V of one million shares per day, and volatility σ of 2 per cent per square-root-of-a-day. Using the invariant price impact constant \$2000 per bet, equation (27) implies that the average price impact cost of a bet satisfies

$$\frac{1}{L} = \left(\frac{2000 \times (0.02)^2}{1^2 \times 40 \times 10^6} \right)^{1/3} = 27 \text{ basis points.} \quad (43)$$

Equations (13), (14), and (15) imply that dollar bet size $E[|PQ|]$, number of bets γ , and the price impact $\lambda Q = \Delta P = S/2$ satisfy

$$\begin{aligned} E[|PQ|] &= CL = 2000 \times 1/0.0027 \approx \$740\,000 \text{ per bet,} \\ \gamma &= \frac{\sigma^2 L^2}{m^2} = \frac{(0.02)^2 \times 0.0027^{-2}}{1^2} = 63 \text{ bets per day,} \\ \frac{1}{2} \frac{S}{P} &= \frac{1}{L} = 27 \text{ basis points.} \end{aligned} \quad (44)$$

Each bet is a block trade of size $Q = \pm \bar{Q} = \pm 18\,500$ shares. While these predictions do not match the empirical literature exactly, these back-of-the-envelope calculations are empirically reasonable, and the predictions are empirically testable.

To summarize, the Treynor model has two nice properties. First, it shows conceptually how price impact is related to adverse selection. Second, the invariance hypothesis $C = \$2000$ per bet makes the Treynor model empirically operational in a very simple manner.

The Treynor model is obviously oversimplified in that it assumes all bets are the same size. This oversimplification leads to two disadvantages. First, it does not allow the price impact of a bet to depend on the bet's size. Second, when bets have different sizes, the moment ratios m and m_β are different from one, and this changes the empirical implications (44) derived from the meta-model (10).

7 The Model of Kyle (1985)

The one-period model of Kyle (1985) fixes both shortcomings of the Treynor model using a model which derives endogenously how the price impact of bets depends linearly on their size.

In the model, a large strategic informed trader has monopolistic access to an informative signal, noise traders trade randomly, and risk neutral market makers trade at break-even prices. Here we describe a more general version of the model, following Bagnoli, Viswanathan and Holden (2001), which does not depend on normal distributions and which allows market makers to observe both the quantity traded by the informed trader and the quantity traded by noise traders without being able to distinguish between the two.

Let Z_F and Z_U denote identically and independently distributed random variables with means of zero and variances of one, not necessarily normally distributed. Market makers believe that the liquidation value satisfies $F = P_0 + \sigma_F Z_F$. An informed trader privately observes the liquidation value F and trades a quantity of bet size $Q_I = x(F)$ shares. A noise trader trades a quantity of bet size $Q_U = \sigma_U Z_U$. Assume that market makers observe the size of both bets $\{Q_I, Q_U\}$, but they do not know which of the two quantities is Q_I and which is Q_U . They set the price $P = p(Q_I, Q_U)$, where p is a symmetric function of its two arguments.¹⁵

An equilibrium is defined as a trading strategy $x(\cdot)$ defining $Q_I = x(F)$ and pricing rule $p(\cdot)$ defining $P = p(Q_I, Q_U)$ such that the informed trader maximizes expected trading profits,

$$x(F) = \operatorname{argmax}_{Q_I} E[(F - p(Q_I, Q_U)) Q_I \mid F], \quad (45)$$

and the market makers, after observing both bets in random order, set break-even prices satisfying the conditional expectation

$$p(Q_I, Q_U) = E[F \mid \{Q_I, Q_U\}]. \quad (46)$$

It is a straightforward exercise to characterize the unique equilibrium with a linear pricing rule $p(Y) = P_0 + \lambda(Q_I + Q_U)$.¹⁶ The assumption that market makers cannot distinguish the informed order from noise trading requires that the quantities Q_I and Q_U appear as a sum $Q_I + Q_U$ in the linear function; they must both have the same coefficient λ .

The first-order condition for the informed trader, $F - P_0 - 2\lambda Q_I = 0$, implies $Q_I = \beta(F - P_0)$,

¹⁵This is more general than the one-period model of Kyle (1985), which assumes that market makers observe the sum $Q_I + Q_U$. Since the market maker cannot tell the difference between the informed bet Q_I and the noise bet Q_U , the pricing rule can be thought of as a function $p(Q_I, Q_U)$ satisfying the symmetry condition $p(Q_I, Q_U) = p(Q_U, Q_I)$. For a linear function p , this implies that the linear coefficients on Q_I and Q_U are the same, consistent with observing the sum.

¹⁶The difficult problem of determining whether equilibria with nonlinear strategies exist is beyond the scope of this paper.

with $\beta = 1/(2\lambda)$. A linear pricing rule $p(Y) = E[F | F/(2\lambda) + U = Y]$ can support an equilibrium only if the conditional expectation is the best linear estimate. Therefore, the least squares coefficient must satisfy the fixed point

$$\lambda = \frac{\frac{\sigma_F^2}{2\lambda}}{\frac{\sigma_F^2}{(2\lambda)^2} + \sigma_U^2}. \quad (47)$$

It immediately follows that the unique candidate equilibrium is

$$Q_I = x(F) = \beta(F - P_0), \quad p(Q_I, Q_U) = P_0 + \lambda(Q_I + Q_U), \quad \text{with } \beta = 1 \frac{\sigma_U}{\sigma_F} \text{ and } \lambda = \frac{1}{2} \frac{\sigma_F}{\sigma_U}. \quad (48)$$

This candidate solution is an equilibrium if and only if the best linear estimate corresponds to the required conditional expectation (46). To show this, observe that the solution implies $Q_I = \sigma_U Z_F$ and $Q_U = \sigma_U Z_U$. Since Z_F and Z_U are identically and independently distributed, the random variables Q_I and Q_U , which scale Z_F and Z_U by the same constant σ_U , are also identically and independently distributed. Even when Z_F and Z_U are not normally distributed, symmetry implies that the least squares solution is the conditional expectation: $E[Z_F | Z_F + Z_U = Y] = Y/2$. The solution $p(Q_I + Q_U)$ in equation (48) simply rescales this result using σ_F and σ_U : $E[F | \sigma_U Z_F + \sigma_U Z_U = Y] = P_0 + (\sigma_F/\sigma_U) Y/2$. This proves equation (48).

The model's predictions (48) are hardly testable in practice "as is", because σ_F and σ_U are conceptually hard to interpret and map into real data. Yet, like the Treynor model, this model is also perfectly consistent with a linear-impact version of the meta-model equations (10), and therefore its predictions can be made empirically operational.

As a result of adverse selection, a random bet $Q = Q_I$ or $Q = Q_U$ moves prices by $\Delta P = \lambda Q$; all of the price impact represents new information about fundamental value. If Z_I and Z_U are ± 1 with equal probabilities, the model is essentially identical to the Treynor model in which all bets are the same size. Unlike the Treynor model, this model otherwise shows why large bets have larger price impact than smaller ones. The model assumption that bets arrive in pairs, one informed and one informationless, should not constrain empirical interpretation.

This model is consistent with the system (10) under the assumption of linear market impact:

$$V = \gamma E[|Q|] \quad (\text{Definition of volume}) \quad (49)$$

$$\sigma^2 = \gamma E\left[\left(\frac{\Delta P}{P}\right)^2\right] \quad (\text{Bets generate all volatility}), \quad (50)$$

$$E[\Delta P^2] = \lambda^2 E[Q^2] \quad (\text{Market impact of one bet}), \quad (51)$$

$$C = \lambda E[Q^2] \quad (\text{Dollar impact cost of a bet.}) \quad (52)$$

$$(53)$$

Mathematically, this model differs from the Treynor model in that the moment ratios m and m_β are not exactly one. Linearity implies $m = m_\beta$, since $\beta = 1$, and the Hölder inequality implies $0 < m \leq 1$. When Z_F and Z_U are normally distributed, we have $m = \sqrt{2/\pi} \approx 0.7979$. Since the model allows Z_F and Z_U to have zero-mean distributions of arbitrary shape, empirical interpretation can be based on empirical evidence about the shape of the distribution of bet size.

It can be checked that all equations in the solution (48) are consistent with the following interpretation of parameters of the model,

$$\sigma_F := \frac{2mP}{L} = \frac{2P\sigma}{\sqrt{\gamma}}, \quad \text{and} \quad \sigma_U := \frac{CL}{mP} = \frac{E[|Q|]}{m}. \quad (54)$$

A reasonable empirical interpretation of this model is that it describes approximately what happens in a market over a period of time in which two bets arrive. The horizon over which the two bets arrive varies across market because the “business clock” runs with different rates of bet arrival γ . The more frequently bets arrive, the more liquid is the market, and the shorter is the business time $1/\gamma$ modeled. Under this interpretation, $\sigma_F/2 = P\sigma/\gamma^{1/2}$ defines dollar volatility in business time as the standard deviation of the price impact of a bet in dollars per share.

As a function of σ_F and σ_U , the cost of a bet C satisfies

$$C = \frac{\sigma_F \sigma_U}{2}. \quad (55)$$

The dollar value C is also equal to the expected profit of the informed trader:

$$E[(F - p(Q_I, Q_U)) Q_I] = \frac{\sigma_F \sigma_U}{2} = C. \quad (56)$$

Thus, invariance of the dollar expected price impact cost of a bet is equivalent to invariance of the dollar expected trading profits of an informed bet, net of market impact costs. The hypothesis that C is invariant can be justified by the hypothesis that informed traders must acquire costly signals in order to earn speculative trading profits. The rate at which informed traders acquire costly signals adjusts endogenously in response to the profitability of doing so. In equilibrium, the informed traders earns zero profits, net of price impact costs and costs of signals. Thus, the invariance hypothesis ultimately is based on the deeper hypothesis that informative signals have the same cost in all markets. The model describes a period of business time during which one informed bet and one noise bet are expected to arrive. We can think of the market as a mechanism which pays for information to be incorporated into prices via informed trading. Under the invariance hypothesis $C = \$2000$, a market with 100 bets per day generates 50 informed bets; altogether, informed traders make \$100 000 per day in trading profits, and noise traders lose \$100 000 per day. The \$100 000 per day compensates the informed traders for the

cost of information.

Numerical Example. Suppose an empirical researcher observes a price P of \$40.00 per share, volume V of one million shares per day, and volatility σ of 2 per cent per square-root-of-a-day. Using the invariant price impact constant of $C = \$2000$ per bet and $m = 0.7979$ given the assumption of normally distributed bet sizes, equation (27) implies that the average price impact cost of a bet satisfies

$$\frac{1}{L} = \left(\frac{2000 \times (0.02)^2}{0.7979^2 \times 40 \times 10^6} \right)^{1/3} = 32 \text{ basis points.} \quad (57)$$

Equations (28), (29), and (30) imply that bet size $E[|Q|]$, number of bets γ , and the price impact coefficient λ satisfy

$$\begin{aligned} E[|Q|] &= \frac{C L}{P} = \frac{2000}{40} \times \frac{1}{0.0032} \approx 15\,625 \text{ shares,} \\ \gamma &= \frac{\sigma^2 L^2}{m^2} = \frac{(0.02)^2 \times 0.0032^{-2}}{0.7979^2} \approx 62 \text{ bets per day,} \\ \lambda &= \frac{m P^2}{C L^2} \approx 62\,335 \text{ dollars-per-shares-squared.} \end{aligned} \quad (58)$$

This example illustrates that the model also generates reasonable empirical predictions.

8 Adverse Selection in the Meta-Model

The meta-model describes how orders affect prices, but it does not relate prices to information or adverse selection. Adverse selection can be easily added to the meta-model, in manner consistent with the Treynor model and the two-bet model, by assuming bets are based on signals which may be imperfectly precise. This requires adding additional invariant parameters measuring the precision of the signals behind informed bets, denoted τ , and the probability a bet is informative, denoted θ . Kyle and Obizhaeva (2017*b*) analyzes this connection formally in the context of a dynamic model; here we summarize some general features of the model, which are likely to be inherent to other models as well.

Let Σ denote the percentage error variance of prices as an estimate of fundamental value, conditional on past public information about prices and quantities:

$$\Sigma = \text{Var} \left[\ln \left(\frac{F}{P} \right) \right]. \quad (59)$$

Then $\Sigma^{1/2}$ has the interpretation of the percentage standard deviation of prices from fundamental value. If $\Sigma^{1/2}$ decreases, prices become more accurate estimates of fundamental value.

This is the measure of market efficiency in the sense of Black (1986); Black believed that markets differ based on how far prices tend to deviate from unobservable fundamental value. Black 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.

This concept of market efficiency differs from Fama’s concept, according to which a market is efficient if all available information is appropriately reflected in price. Fama’s concept of efficiency implies that prices—adjusted for the risk-free rate, dividend yield, and risk premium—follow a martingale regardless of how much information is available overall in the market. Here we assume that markets are efficient according to Fama’s definition and use this assumption to model the degree of efficiency according to Black’s definition.

Consider a signal which has precision τ in the sense that, if publicly announced, prices would react in such a way that error variance Σ would be reduced by factor τ to $(1 - \tau)\Sigma$. If an informed trader trades on the signal, it reduces error variance by a factor less than τ because the market cannot distinguish informed signal from uninformed ones. If a given bet has precision τ with probability θ and precision zero with probability $1 - \theta$, then—from the perspective of market makers—it can be shown mathematically that each bet reduces error variance Σ to approximately $(1 - \theta^2\tau)\Sigma$.¹⁷

Since each of γ bets per day contributes σ^2/γ to returns variance, we obtain

$$\frac{\sigma^2}{\gamma} := E\left[\left(\frac{\Delta P}{P}\right)^2\right] = \frac{\lambda^2 E[Q^2]}{P^2} = \theta^2\tau\Sigma. \quad (60)$$

Plugging equation (29) for γ makes it possible to express Σ as a function of either L or γ :

$$\Sigma = \frac{\sigma^2}{\gamma(\theta^2\tau)} = \frac{m^2}{(\theta^2\tau)L^2}. \quad (61)$$

Now assume that the parameters θ and τ are both invariant. Then equation (61) has a remarkable interpretation: By observing the level of liquidity L , a trader can infer the error variance of the price Σ from knowledge of the invariant parameters θ and τ . When liquidity L is high, the average distance between prices and fundamentals $\Sigma^{1/2}$ is low.

Equation (61) is consistent with the Treynor model ($m = 1$) and the two-bet model. These models make the simplistic assumption that the informed trader observes the fundamental

¹⁷If market makers could distinguish informed bets from uninformed bets, then each bet would reduce error variance by approximately $(1 - \theta\tau)\Sigma$, not $(1 - \theta^2\tau)\Sigma$. Intuitively, the parameter θ is squared because anonymous information is deflated by factor θ twice, once because a given bet may or may not be informed and once again because the informed bets are indistinguishable from the noise bets. In the two-bet model, $\theta = 1/2$ and $\tau = 1$; since two bets reduce Σ by a factor $1/2$, each bet reduces error variance by factor $\theta^2 = 1/4$ from the perspective of the market makers, who cannot distinguish informed from uninformed bets.

value exactly ($\tau = 1$). The error variances can be expressed in terms of deep parameters such as ΔF and σ_F and then related to liquidity measure L using models' solutions in the following manner:

$$\text{Treynor model:} \quad \Sigma = \left(\frac{\Delta F}{P} \right)^2 = \frac{1}{\theta^2 L^2}, \quad (62)$$

$$\text{Kyle model:} \quad \Sigma = \frac{\sigma_F^2}{P^2} = \frac{4m^2}{L^2}. \quad (63)$$

These equations are identical to equation (61), because $m = 1$ in the Treynor model and $\theta = 1/2$ in the modified Kyle model, in both models $\tau = 1$.

The concept of pricing accuracy is closely related to the concept of market resiliency. Let market resiliency $\rho(t)$ denote the rate at which past pricing errors decay. The higher is the rate, the faster the market recovers after different shocks. Trading by informed traders and release of public information push prices towards fundamentals. Assume that private signals and public information are noisy observations of the pricing error $F - P$. Then it can be shown mathematically that since each bet reduces error variance Σ by the factor $\theta^2 \tau$ and there are on average γ bets per day, resiliency ρ is equal to

$$\rho = (\theta^2 \tau) \gamma. \quad (64)$$

The percentage pricing error $(F - P)/P$ is approximately an AR-1 process with mean reversion rate ρ . Equation (61) yields that the resilience is related to pricing accuracy Σ and liquidity L by

$$\rho = \frac{\sigma^2}{\Sigma} = \frac{\theta^2 \tau}{m^2} \sigma^2 L^2. \quad (65)$$

Resiliency is high when pricing accuracy $\Sigma^{-1/2}$ is high and market liquidity L is substantial.

To summarize, by introducing the invariant quantity $\theta^2 \tau$ measuring the precision of a bet, we have mapped the meta-model into the concept of adverse selection. We have obtained hard-to-measure values of pricing error variance Σ and market resiliency ρ as simple functions of easy-to-measure parameters and other invariant quantities. The adverse-selection related to quantity $\theta^2 \tau$ does not show up in the meta-model equations (10) but does show up in the equations for pricing error and resiliency. The value of Σ is proportional to $1/L$, and the value of ρ is proportional to γ .

For example, suppose that $\theta = 1/2$, $\tau^{1/2} = 0.01$, $m = 0.25$, $\sigma = 0.02/\text{day}^{1/2}$, and $1/L = 0.0040 = 40$ basis points. Then $\gamma = 400$ bets/day, $\Sigma^{1/2} = 0.20 = 20\%$, and $\rho = 100$ days.

9 Smooth Trading Model

In the block-trading models discussed previously, a trader is allowed to trade only once. While the model of Kyle, Obizhaeva and Wang (2018) has a near-closed-form solution, the equations describing this solution are quite complicated. Here we discuss a simpler model in which traders receive private information only once, but trade on a given signal gradually over time. This model is a natural generalization of the block trading model, which describes trading in the late 20th century, to a model of flow trading, which describes trading in the 21st century.

From the perspective of an individual trader, the price is a function of all other traders' information, the trader's own inventory $S(t)$, and the the time derivative of the trader's inventory:

$$P(t) = P_{-*}(t) + \lambda S(t) + \kappa x(t), \quad \text{where} \quad x(t) = \frac{dS(t)}{dt}. \quad (66)$$

Here, λ measures permanent price impact in the same way as the two-bet model. The parameter κ measures temporary price impact because the effect on prices of trading at rate $x(t)$ immediately goes away when the trader stops trading. The units for λ (dollars/shares²) are the same as before (with the linearity assumption $\beta = 1$). The units for κ (dollars \times days/shares²) are the units for λ multiplied by days.

To simplify this model, we assume that permanent price impact λ is so small that it can be ignored. We therefore focus mostly on temporary price impact. This is consistent with the intuition that traders arriving into the market are risk neutral. Therefore, the price is given by

$$P(t) = P_{-*}(t) + \kappa x(t), \quad \text{where} \quad x(t) = \frac{dS(t)}{dt}. \quad (67)$$

Suppose a trader obtains a signal about the difference between unobservable fundamental value F and observable market price P . He believes that the estimate of $F - P$ has to be adjusted by ΔF . As time passes, other traders learn about the same information and, if a trader does not trade, the price evolves as

$$P_{-*}(t+h) = P(t) + \Delta F \left(1 - e^{-\rho h}\right) \quad (\text{No-Trade Price}). \quad (68)$$

In the distant future $h \rightarrow \infty$, all information is incorporated into prices, even if the trader does not trade, because of learning and trading by other traders. The difference between the trader's estimate of the fundamental value and this no-trade price $P_{-*}(t+h)$ at time $t+h$ is $\Delta F e^{-\rho h}$; trader's private information ΔF decays at a rate ρ , the rate at which prices adjust. Asset managers refer to $\Delta F/P$ as their "alpha" and refer to idea that the signal eventually is incorporated into prices without their own trading as "alpha decay."

If the informed trader trades, then he chooses to incorporate a fraction θ of his private information into prices by trading initially trading $\Delta x(t) = \theta \Delta F / \kappa$ at time t and subsequently $\Delta x(t+h) = \Delta x(t) e^{-\rho h} / \kappa$ at time $t+h$. The trader believes the price at time $t+h$ is given in expectation by

$$E[P(t+h) | \Delta F, x(t+h)] = P(t) + \Delta F \left(1 - e^{-\rho h}\right) + \theta \Delta F e^{-\rho h} \quad (\text{Informed Price}). \quad (69)$$

In this equation, the first term involving ΔF represents the effect of other traders expected learning (alpha decay) and the second term represents the traders own price impact. If the trader is actually informed, then he makes a profit, correctly anticipating that the price will eventually change by $\Delta F(t)$. The informed trader's expected profit is

$$E[\text{Informed Trader Profit} | \Delta F] = \frac{\theta(1-\theta) \Delta F^2}{2\kappa \rho}. \quad (70)$$

Choosing $\theta = 1/2$ maximizes a risk neutral trader's expected profit.

If the trader believes himself to be informed but he is actually uninformed, then the price does not change as result of other traders learning; we have $P_{-*}(t+h) = P(t)$ with correct expectations. Since the uninformed trader follows the same trading strategy as the informed trader, the price dynamics after his trade is

$$P(t+h) = P(t) + \theta \Delta F e^{-\rho h} \quad (\text{Uninformed Price}). \quad (71)$$

In this equation, there is no effect from other traders learning and the term involving ΔF represents the traders own price impact from trading on an uninformative signal. The uninformed trader loses money from price impact by incorrectly trading on a forecast that the long-term price will change by $\Delta F_U(t)$, when in fact the long-term price does not change at all because his signal generates no alpha. His expected loss is

$$E[\text{Uninformed Trader Profit} | \Delta F] = -\frac{\theta^2 \Delta F^2}{2\kappa \rho}. \quad (72)$$

Now suppose that market makers believe that the trader is informed with probability θ and uninformed with probability $1 - \theta$. The model requires the probability that the trader is informed to equal the fraction of information that the trader incorporates into price.¹⁸ The intuition for this is the following: If the market makers believe that a fraction less than θ are in-

¹⁸This approach involving informed and uninformed traders is consistent with the dynamic model in Kyle and Obizhaeva (2017b). The model of Kyle, Obizhaeva and Wang (2018) assumes instead that traders are relatively overconfident.

formed, they will offer greater liquidity; this will make informed trading more profitable and more informed traders will enter by purchasing costly information, eventually making the fraction informed greater than θ . If the market makers believe the fraction of informed traders is greater than θ , then market makers will reduce liquidity toward zero, making informed trading unprofitable; informed traders will not acquire costly information and the fraction informed will drop below θ . When the fraction of informed traders is exactly θ , there is an equilibrium in which the informed traders provide just enough market depth so that trading on private information covers the cost of acquiring it.

Since the market makers believe the fraction of informed traders is θ , they set the price equal to a weighted average of the informed price (69) with probability weight θ and uninformed price (71) with probability weight $1 - \theta$. This results in the price jumping by $\theta \Delta F$ at time t to

$$E[P(t+h) \mid \text{Bet arrives}] = P(t) + \theta \Delta F. \quad (73)$$

The price follows a martingale, consistent with Fama's concept of market efficiency, because risk neutral market makers extract information from the order flow correctly.

The total quantity traded by an informed or uninformed traders is can be mapped into a bet as

$$Q(t) = \int_{h=t}^{\infty} \Delta x(t) e^{-\rho h} dh = \frac{\Delta x(t)}{\rho}, \quad (74)$$

where $\Delta x(t) = \theta \Delta F / \kappa$ denotes the immediate change in the rate of trading upon arrival of new information. Given that the rate of bet arrival is $\gamma(t)$, the expected bet volume is

$$V = \gamma \frac{E[|\Delta x|]}{\rho}. \quad (75)$$

The price impact of a new bet is given by

$$E[P(t+h) - P(t) \mid \text{Bet Arrives}] = \kappa \Delta x(t). \quad (76)$$

The entire jump at time t is due to the temporary price impact of the trader beginning to trade at a rate $\Delta x(t)$. For time $t+h > t$, the price impact takes into account both the declining temporary price impact of the trading itself and the increasing learning by other traders (alpha decay). Thus, the market impact of each bet is

$$E[(\Delta P)^2] = \kappa^2 (\Delta x(t))^2. \quad (77)$$

The total cost of the bet, including both temporary price impact and alpha decay, from trad-

ing $Q(t) = \Delta x(t)/\rho$ with impact $\kappa \Delta x(t)$ is

$$C = \frac{\kappa (\Delta x(t))^2}{\rho}. \quad (78)$$

This model of smooth trading satisfies the following meta-model equations:

$$V = \gamma \frac{E[|\Delta x|]}{\rho} \quad (\text{Definition of volume}) \quad (79)$$

$$\sigma^2 = \gamma E \left[\left(\frac{\Delta P}{P} \right)^2 \right] \quad (\text{Bets generate all volatility}), \quad (80)$$

$$E[(\Delta P)^2] = \kappa^2 (\Delta x)^2 \quad (\text{Market impact of one bet}), \quad (81)$$

$$C = \kappa \frac{(\Delta x)^2}{\rho} \quad (\text{Dollar impact cost of a bet.}) \quad (82)$$

This system is identical to meta-model system (10) for block trading if one makes the following assumptions about bet sizes and

$$Q \mapsto \frac{|\Delta x|}{\rho} \quad (\text{Bet Size}) \quad (83)$$

$$\lambda \mapsto \kappa \rho \quad (\text{Market Impact}) \quad (84)$$

In the smooth trading model, a bet corresponds to the total trading volume that a shock to target inventories generate over time (ignoring for simplicity volume generated during liquidation of the position in the distant future). The resiliency parameter ρ maps temporary price impact into permanent price impact and also maps flow trading into intended bets. Since the resiliency parameter incorporates the effect of adverse selection, we have obtained a nice meta-model which directly includes the effect of adverse selection!

Based on the solutions (28) for the block-trading meta-model and the equation for $\rho = (\theta^2 \tau)\gamma$, one can simply obtain solutions for the rate of trading $x(t)$ and temporary price impact κ for the framework of smooth trading in terms of the measure of liquidity L :

$$E[|\Delta x|] = \rho E[|Q|] = \theta^2 \tau V, \quad (85)$$

$$\kappa = \frac{\lambda}{\rho} = \frac{m^2 P}{\theta^2 \tau V L}. \quad (86)$$

Unlike the block trading model, both the rate of trading associated with a bet and price impact depend on the information content of bets $\theta^2 \tau$.

When written so that the left and right sides are dimensionless, these equations are equiva-

lent to

$$E \left[\frac{|\Delta x|}{V} \right] = \theta^2 \tau, \quad (87)$$

$$\frac{\kappa E[|\Delta x|]}{P} = \frac{m^2}{L}. \quad (88)$$

The first equation gives that rate of initial buying of a bet as a fraction of one day's volume; a trader should buy at a rate equal to one percent of a day's volume if the precision of the signal $\theta^2 \tau$ is one percent! The second equation says that the price impact from flow trading immediately after receiving private information has impact m^2/L .

10 Trading Liquidity, Funding Liquidity, and Time

Market liquidity is closely related to time. Markets operate in business time, which differs from calendar time and naturally relates to the speed with which bets arrive into the market, defined equivalently in several ways as

$$\gamma = \left(\frac{PV\sigma}{mC} \right)^{2/3} = \frac{PV}{CL} = \frac{\sigma^2 L^2}{m^2}. \quad (89)$$

A fundamental implication of market microstructure invariance is that less liquid markets operate at a slower rate and are less resilient than more liquid markets. The close connection between liquidity and time has important implications for many areas of finance. It is particularly important in banking, because traditional banks hold illiquid loans on their balance sheets. Here we discuss several issues related to banking.¹⁹

Repo Haircuts and Margins. Consider the appropriate haircut for a repurchase agreement in which a risky asset is used as collateral to finance a short-term loan. The haircut is the difference between the amount borrowed and the size of the loan; it measures the borrower's equity in the risky asset as a fraction of the value of the asset. If the borrower of cash defaults on the loan, the lender of cash takes control of the collateral and sells it. If the proceeds of the sale provide enough cash to make the lender whole, the balance is paid to the borrower. Similar intuition underlies setting margin requirements, which help to protect a clearing houses against defaults of its members.

Traditional thinking about both haircuts and margin requirements is based on the idea that the haircut or margin requirement should be a multiple of the standard deviation of returns

¹⁹Kyle and Obizhaeva (2016b) discuss similar issues in the context of fixed income markets.

on the asset. This intuition makes sense if collateral is marked to market infrequently, at say interval ΔH (days), in which case a haircut equal to a multiple of $\Delta H^{1/2} \sigma$ makes sense. With modern technology, collateral can be valued very quickly and margin calls can be made very quickly, even intraday. This suggests a small value of ΔH and thus a small haircut based on the asset's volatility.

Invariance implies that markets run at different pace. The liquidity of an asset is traditionally defined as the ability to sell the asset quickly at the current market price without loss. Since the collateral may be liquidated over some period of time H (days), to save on transaction costs, repo haircuts and margin requirements have to be proportional to the standard deviation of the gain or loss $\sigma H^{1/2}$. Since the optimal period of collateral liquidation is related to how fast the market work $1/\gamma$, the repo haircuts and margins must be proportional to $\sigma\gamma^{-1/2}$ or, given the equation (89), to m/L . Thus, the liquidity measure $1/L$ is a measure of both trading liquidity and funding liquidity.

Sometimes liquidation of collateral occurs at a rapid pace. It is then often called a “fire sale” because the temporary price impact of liquidation reverses itself after the sales are completed. The smooth trading model is especially attractive for explaining fire sales because it allows temporary price impact. Extremely rapid selling (over short horizon H) can result from a moral hazard problem. If the lender must return to the borrower sales proceeds over and above the amount owed but must suffer the losses when the sales proceeds are insufficient to cover the amount owed, the lender has an incentive to sell collateral more quickly if there is great uncertainty about whether collateral sales will cover the amount owed. The measure of liquidity L may also help to quantify transaction costs involved in fast liquidation of a collateral.

Bank Capital. The logic applied to repo haircuts above also applies to capital requirements for banks. A bank typically holds a portfolio of safe but illiquid assets. The assets are safe in the sense that the portfolio's return standard deviation σ is small. The assets are illiquid in that business time passes very slowly; γ is very small. A traditional bank holds assets which trade infrequently, if at all, and are difficult, if not impossible, to sell off if the bank fails. Its creditors, including the government as lender of last resort and insurer of deposits, have an incentive to make sure that the bank has a capital cushion large enough to avoid taxpayer financed bailouts. The logic for repo haircuts suggests that bank capital should be proportional to $1/L$, where L is the liquidity of the bank's portfolio of assets. Since $1/L = \sigma/(m\gamma^{1/2})$, the optimal capital requirement depends on the relationship between the riskiness of its assets σ and the speed with which they can be liquidated γ .

Bank Equity Issuance. Several specific features of banks give them a tendency to “blow up.” First, bank assets are typically bonds. Investment-grade or equivalent bonds have low credit risk, in the sense that the standard deviation of their returns is small. When the bonds deteriorate in quality, their return standard deviations increase substantially. This increases the return volatility of the individual bonds. Second, banks are highly leveraged firms, with equity in a healthy bank representing equal to say 10% of the banks assets. If the bank’s assets deteriorate in quality, the bank’s capital evaporates quickly and this magnifies the volatility of the return on the bank’s equity. As a result, it becomes very expensive for the bank to issue new equity; the bank has blown up.

Capital requirements should also take into account the fact that the volatility of bank assets increases when the assets deteriorate in quality. This lessens their ability to be sold and makes optimal capital requirements higher. The theoretical model of Kyle and Xiong (2001) shows how changes in bank capital increase volatility through wealth effects even if the fundamental volatility of the asset does not change.

Government Securities. Government securities lie at the opposite end of the spectrum of liquidity from traditional bank assets. Government securities are generally safe because the probability of default is less than the probability of default by private firms. They are actively traded because governments continuously roll over large amounts of debt. The combination of high volume and low volatility makes government securities highly liquid in the sense that L is very large. Of course, traders seeking liquid assets to hold flock to government securities; this increases trading volume and makes them even more liquid.

Stock Market Crashes. While business time passes very slowly in markets for illiquid debt instruments, it passes very fast in the markets for overall stock indexes, such as the S&P E-mini futures contract and index exchange traded funds (ETFs). In these markets, both L and γ are very large. Despite high liquidity, when gigantic selling pressure occurs, stock markets may not be able to absorb selling pressure without large price reaction. Kyle and Obizhaeva (2016c) examine five stock market crashes by comparing the size of actual price declines with theoretical predictions based on liquidity measures (L and γ) and dollar amounts sold. The the stock market crash of 1929, which occurred over several months, had more modest price declines than predicted from a block trading model based on L . The crash of 1987 and the resulting from Societe Generales’s liquidation of Jerome Kerviel’s rogue trades in 2008, which occurred over periods of several days, have magnitudes reasonably consistent with the value of L in a block trading model. Two flash crashes—the 1987 sales by George Soros and the 2010 flash crash—occurred over period of minutes and had have much larger price declines than

predicted by the block trading model relative to the dollar values traded. Since the large declines were immediately followed by rapid reversals, these two flash crashes are consistent with fire sales in the context of a smooth trading model. The fact that price declines in 1929 were smaller than predicted from a block trading model also suggests that the smooth trading model may be relevant for explaining price pressure.

11 Conclusion

Liquidity is a confusing concept because it is hard to pin down precisely what it means and how to measure it. Here we have shown how the invariance-implied liquidity measure L , defined using trading volume and returns volatility, provides a practical, empirical link between theoretical models of liquidity and practical ways to measure it. Using the hypothesis that some variables, such as the market impact cost of a bet, are invariant, the same liquidity measure naturally follows from three different minimalist approaches based on a (1) meta-model, (2) invariance of business time, or (3) dimensional analysis and leverage neutrality.

The variable L measures liquidity in many ways. It describes bid-ask spreads, market impact, the size of bets, and the speed with which business time passes (the rate at which bets arrive). Although the definition of L does not refer to measures of private information or adverse selection, such as the probability of informed trading and precision of private information, theoretical models of market microstructure based on adverse selection map nicely into this metrics as well; this makes it possible to use L to identify pricing accuracy and price resiliency. Since L is also related to the concept of business time, an internal clock with which markets operate, and therefore it may also help to generate important insights about various quantities related to funding liquidity and financial intermediation in general. The concept generalizes nicely from models of block trading to models of smooth trading. The idea that liquidity and information are both intrinsically flow concepts suggests that trading in risk assets should be explicitly structured to allow trading to take place slowly as flows, not quickly as blocks. This can be implemented by structuring flow trading around simple messages that allow large traders to control their rates of buying and selling without having to compete on bandwidth and message costs with high frequency traders

References

Amihud, Yakov. 2002. "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects." *Journal of Financial Markets*, 5: 31–56.

- Andersen, Torben G., Oleg Bondarenko, Albert S. Kyle, and Anna A. Obizhaeva.** 2015. "Intra-day Trading Invariance in the E-Mini S&P 500 Futures Market." Working Paper. Available at <http://dx.doi.org/10.2139/ssrn.2693810>.
- Bae, Kyoung-hun, Albert S. Kyle, Eun Jung Lee, and Anna A. Obizhaeva.** 2016. "Invariance of Buy-Sell Switching Points." Robert H. Smith School Research Paper No. RHS 2730770. Available at Available at SSRN: <https://ssrn.com/abstract=2730770> or <http://dx.doi.org/10.2139/ssrn.2730770>.
- Bagnoli, Mark, S. Viswanathan, and Craig Holden.** 2001. "On the Existence of Linear Equilibria in Models of Market Making." *Mathematical Finance*, 11(1): 1–31.
- Barra.** 1997. *Market Impact Model Handbook*. Berkley, CA: Barra.
- Benzaquen, Michael, Jonathan Donier, and Jean-Philippe Bouchaud.** 2016. "Unravelling the Trading Invariance Hypothesis." Working Paper. Available at <https://arxiv.org/abs/1602.03011>.
- Black, Fischer.** 1971. "Toward a Fully Automated Exchange, Part I." *Financial Analysts Journal*, 27(4): 28–35.
- Black, Fischer.** 1986. "Noise." *Journal of Finance*, 41(3): 529–543.
- Black, Fischer.** 1995. "Equilibrium Exchanges." *Financial Analysts Journal*, 51(3): 23–29.
- Budish, Eric, Peter Cramton, and John Shim.** 2015. "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response." *Quarterly Journal of Economics*, 130(4): 1547–1621.
- Christie, William G., and Paul H. Schultz.** 1994. "Why do NASDAQ Market Makers Avoid Odd-Eighth Quotes?" *Journal of Finance*, 49(5): 1813–1840.
- Diamond, Douglas W., and Robert E. Verrecchia.** 1981. "Information Aggregation in a Noisy Rational Expectations Economy." *Journal of Financial Economics*, 9(3): 221–235.
- Grinold, Richard C., and Ronald N. Kahn.** 1995. *Active Portfolio Management*. Probus Publishing.
- 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." *Journal of Finance*, 43(3): 617–633.

- Harris, Lawrence.** 2015. "Transaction Costs, Trade Throughs, and Riskless Principal Trading in Corporate Bond Markets." Working Paper. Available at SSRN: <https://ssrn.com/abstract=2661801> or <http://dx.doi.org/10.2139/ssrn.2661801>.
- Kirilenko, Andrei, Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun.** 2017. "The Flash Crash: High-Frequency Trading in an Electronic Market." *Journal of Finance*, 72(3): 967–998.
- 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(3): 317–355.
- Kyle, Albert S., and Anna A. Obizhaeva.** 2016a. "Market Microstructure Invariance: Empirical Hypotheses." *Econometrica*, 84(4): 1345–1404.
- Kyle, Albert S., and Anna A. Obizhaeva.** 2016b. "Trading Liquidity and Funding Liquidity in Fixed Income Markets: Implications of Market Microstructure Invariance." Working Paper. Available at <https://www.frbatlanta.org/-/media/documents/news/conferences/2016/0501-financial-markets-conference/papers/kyle-trading-liquidity-and-funding-liquidity.pdf>.
- Kyle, Albert S., and Anna A. Obizhaeva.** 2017a. "Dimensional Analysis, Leverage Neutrality, and Market Microstructure Invariance." Working Paper. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.2785559>.
- Kyle, Albert S., and Anna A. Obizhaeva.** 2017b. "Market Microstructure Invariance: A Dynamic Equilibrium Model." Working Paper. Available at Available at SSRN: <https://ssrn.com/abstract=2749531> or <http://dx.doi.org/10.2139/ssrn.2749531>.
- Kyle, Albert S., and Anna A. Obizhaeva.** 2018. "The Market Impact Puzzle." Working Paper. Available at SSRN: <https://ssrn.com/abstract=3124502> or <http://dx.doi.org/10.2139/ssrn.3124502>.
- Kyle, Albert S., and Anna Obizhaeva.** 2016c. "Large Bets and Stock Market Crashes." Working Paper. Available at SSRN: <https://ssrn.com/abstract=2023776> or <http://dx.doi.org/10.2139/ssrn.2023776>.
- Kyle, Albert S., and F. Albert Wang.** 1997. "Speculation Duopoly with Agreement to Disagree: Can Overconfidence Survive the Market Test?" *Journal of Finance*, 52(5): 2073–2090.

- Kyle, Albert S., and Jeongmin Lee.** 2017. "Toward a Fully Continuous Exchange." *Oxford Review of Economic Policy*, 33(4): 650–675.
- Kyle, Albert S., and Wei Xiong.** 2001. "Contagion as a Wealth Effect." *Journal of Finance*, 56(4): 1401–1440.
- Kyle, Albert S., Anna A. Obizhaeva, and Tugkan Tuzun.** 2017. "Microstructure Invariance in U.S. Stock Market Trades." Working Paper. Available at SSRN: <https://ssrn.com/abstract=1107875> or <http://dx.doi.org/10.2139/ssrn.1107875>.
- Kyle, Albert S., Anna A. Obizhaeva, and Yajun Wang.** 2017. "Beliefs Aggregation and Return Predictability." Working Paper. Available at SSRN: <https://ssrn.com/abstract=2639231> or <http://dx.doi.org/10.2139/ssrn.2639231>.
- Kyle, Albert S., Anna A. Obizhaeva, and Yajun Wang.** 2018. "Smooth Trading with Overconfidence and Market Power." *Review of Economic Studies*, 85: 611–662.
- Kyle, Albert S., Anna A. Obizhaeva, Nitish Sinha, and Tugkan Tuzun.** 2010. "News Articles and Equity Trading." Working Paper. Available at <http://dx.doi.org/10.2139/ssrn.1786124>.
- Lee, Jeongmin, and Albert S. Kyle.** 2017. "Information and Competition with Speculation and Hedging." Working Paper. Available at SSRN: <https://ssrn.com/abstract=2892141> or <http://dx.doi.org/10.2139/ssrn.2892141>.
- Lee, Jeongmin, and Albert S. Kyle.** 2018. "When are Financial Market Perfectly Competitive?" Working Paper.
- Lindsey, James D., Charles D. Hutton, Joe W. Tubb, Carol L. Shipman, and Albert S. Kyle III.** 1991. "Goods Database Employing Electronic Title or Documentary-type Title." US Patent 5,063,507.
- Lindsey, James D., Charles D. Hutton, Joe W. Tubb, Carol L. Shipman, and Albert S. Kyle III.** 1994. "Method for Carrying out Transactions of Goods Using Electronic Title." US Patent 5,285,383.
- Treynor, Jack.** 1995. "The Only Game in Town." *Financial Analysts Journal*, 51(1): 81–83 (reprint of Walter Bagehot, pseud. 1971. "The Only Game in Town." *Financial Analysts Journal*, 22(2): 12–14, 22).
- Vayanos, Dimitri.** 1999. "Strategic Trading and Welfare in a Dynamic Market." *Review of Economic Studies*, 66(2): 219–254.