THE MARKET IMPACT PUZZLE

Albert S. Kyle
Anna A. Obizhaeva

Working Paper
No 270

NES Working
Paper series

August
2020

The Market Impact Puzzle

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

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

First Draft: February 4, 2018

Abstract

Finding a universal market impact formula remains one of the most fascinating puzzles in finance. This paper reviews two possible approaches for imposing restrictions on this formula. First, restrictions can be obtained from a system of economic equations using trading volume and volatility, as suggested by Kyle and Obizhaeva (2017*b*). Second, restrictions can be derived using dimensional analysis and leverage neutrality, as suggested by Kyle and Obizhaeva (2017*a*). Except for the knife-edged case of the square root market impact function, additional assumptions related to market microstructure invariance are needed to apply the same market impact formula to all assets simultaneously. This results in a tightly parameterized universal market impact formula suitable for empirical testing.

Keywords: Market microstructure, invariance, liquidity, square root model, market impact, transaction costs, dimensional analysis, leverage neutrality, volume, volatility.

*Albert S. Kyle has worked as a consultant for various companies, exchanges, and government agencies. He is a non-executive director of a U.S.-based asset management company.

1 Introduction

Understanding market impact is one of the most puzzling and difficult issues in finance. A market impact function describes how much prices are expected to move away from pre-trade benchmarks upon execution of a trade.

For the purpose of managing transaction costs, it is essential for an asset manager to have an operational formula for market impact that is easy to use. Since different strands of financial markets research generate different implications for market impact, research on market impact is central to our understanding of how financial markets work.

Despite many decades of theoretical and empirical research, there are still many open questions concerning market impact. Financial assets have diverse characteristics and different trading patterns. There are large common stocks, small common stocks, corporate bonds, sovereign debt, foreign currencies, commodities, and options. Some of these are traded on electronic exchanges with high frequency trading; others are traded in dealer markets by telephone. Is it necessary to estimate market impact functions for each asset or asset type separately? Or does a single magic formula describe transaction costs for dissimilar assets simultaneously? If a universal market impact exists, what is its functional form, and how does it relate to various asset characteristics? Market impact is also a puzzle because finance theory suggests a market impact function has a linear functional form, with price impact proportional to the number of shares traded, while empirical evidence suggests a square root model, with marginal price impact diminishing as the number of shares traded increases.

The purpose of our paper is to use theory to impose discipline on solving this puzzle by restricting the functional form of the universal market impact function so that it includes both square root and linear models as special cases, while simultaneously having as few degrees of freedom as possible.

The *square root model* of market impact was proposed by Barra (1997), based on empirical regularities observed by Loeb (1983), as a practical way for asset managers to measure market impact empirically. Often referred to as “the Barra model,” it is discussed in the well-known handbook for asset managers by Grinold and Kahn (1995).

The square root model is strikingly simple. Let G denote the percentage cost of executing a bet or meta-order of Q shares of stock with price P, expressed as a fraction of the value of the bet $|PQ|$. Let σ^2 denote the asset's returns variance per day, and let V denote the asset's trading volume in shares per day. Then the square root model of market impact can be written

$$G = g(\sigma, P, V; Q) \quad \sim \quad \sigma \left(\frac{|Q|}{V}\right)^{1/2}, \tag{1}$$

where $g(\ldots)$ provides a functional form for the model and the notation “$\sim$” means “is proportional to.” Trading volume and returns volatility are easy to observe or estimate. Empirical estimates suggest that the proportionality constant is close to one, and theory implies that it is exactly one for each asset; thus, the same model can be applied to all assets simultaneously. For example, if daily volatility is 2 percent and trading volume is one million shares per day, then execution of 5 percent of daily trading volume is expected to move prices or to cost a trader about 45 basis points $(0.02\,(0.05)^{1/2})$, or about 0.45% of the dollar value traded.

The square root formula is dimensionally consistent. Bet size Q is measured in shares, volume V is measured in shares per day, and returns variance σ^2 is measured per day (implying returns standard deviation or volatility is measured per square-root-of-a-day). The proportionality coefficient in (1) is dimensionless. Market impact G a dimensionless quantity, which is the same regardless of the units of measurement of Q, V, and σ^2.

This square root model is elegantly simple and empirically realistic. Yet, it may be a bit too simple and not quite empirically realistic enough. The square root model has two distinct disadvantages.

First, even though the square root model seems to be a reasonable approximation to empirical estimates of transaction costs, there is still no consensus on whether market impact functions can indeed be described exactly by the square root function. The econophysics literature often finds supportive evidence for the square-root price-impact function, as discussed by Gabaix et al. (2006). Some empirical studies, like Almgren et al. (2005) and Kyle and Obizhaeva (2016), find an exponent closer to 0.60 than the exponent of $1/2$. Empirically, it is even difficult to distinguish between a square root function and a linear function with proportional bid-ask spread.

Market impact models are difficult to estimate empirically. As discussed by Obizhaeva (2012), data samples are usually subject to severe selection bias because decisions to execute trades are determined endogenously and often depend on contemporaneous price dynamics, making market impact estimates biased. Traders may cancel the balance of partial executions of very expensive orders and execute cheap orders fully. As a result, liquidity estimates based on executed transactions are often biased downwards because the opportunity cost of canceled orders is not taken into account. Moreover, since implementation shortfall is very noisy, empirical tests have low statistical power; therefore, many observations are needed to obtain reasonably precise estimates of trading costs. Furthermore, controlled experiments are difficult or costly.

Second, the square-root model is hard to reconcile with theoretical research. Following the methodology proposed by Kyle and Obizhaeva (2017*a*), Pohl et al. (2017) point out that volume and volatility are sufficient for deriving a square root model using dimensional analysis and leverage neutrality. By contrast, most theoretical models of market microstructure lead to a model of linear market impact, not a square root model. Many theoretical models, such as Kyle (1985), obtain endogenously linear market impact by combining normally distributed random variables describing information about assets' payoffs with either exponential utility or risk neutrality characterizing traders' preferences. Models with non-linear market impact are usually analytically intractable and seem to allow for simple arbitrage strategies. If market impact is sufficiently concave in the size of bets, then one could hypothetically make profits simply by making many purchases in small tranches and then selling the entire accumulated position as one large order. For example, given market impact (1), one could make profits by executing over time ten buy trades of 100 shares each and then selling 1000 shares at once.

Proposing principles of *market microstructure invariance*, Kyle and Obizhaeva (2016) obtain a more general functional form for a universal market impact function, derived from the intuition that all markets operate similarly, except for operating at different speeds related to the pace of arrival of new bets (or trading ideas). Liquid assets are traded in fast markets, and illiquid assets are traded in slow markets. However dissimilar markets may seem to be in calendar time, all markets are the same in business time in a sense that (1) dollar risks transferred by bets in business time and (2) the dollar costs of executing bets transferring comparable risks are approximately the same across all markets.

These invariance hypotheses imply the empirical conjecture that market impact has the more general functional form

$$G = g(\sigma, P, V; Q) \quad \sim \quad \left(\frac{\sigma^2}{PV}\right)^{1/3} f\left(\left(\frac{\sigma^2}{PV}\right)^{1/3} |PQ|\right). \tag{2}$$

This formula is empirically and theoretically attractive because the special case $f(Z) \sim |Z|^{\beta}$ nests both the square root model with exponent $\beta = 1/2$ and the linear model with exponent $\beta = 1$. Furthermore, the functional form f nests an "invariance hypothesis" which implies that the dimensional proportionality coefficient is the same for all assets. Kyle, Obizhaeva and Kritzman (2016) illustrate the main results using some numerical examples.

Although more general than the square root model, this formula is still tightly parameterized. Given the dollar size of a bet $|PQ|$, the only asset characteristic needed to measure market impact is the ratio of returns variance to dollar volume $\sigma^2/(PV)$.

In this paper, we review two different theoretical approaches for deriving a generalized market impact formula. These two approaches are based on general theoretical principles that impose some restrictions on a universal market impact function. Both approaches converge to the same outcome, a tightly parameterized general market impact formula (2) with the proportionality constant predicted to be the same or almost the same for all assets, regardless of how different these assets are.

The first *"meta-model"* approach imposes intuitive economic restrictions that are likely to be consistent with many theoretical models. These restrictions describe how bets (or metaorders) add up to volume, generate market impact, and induce price volatility. Another restriction based on market microstructure invariance imposes an additional empirical relationship on the size and number of bets. These restrictions are combined into a system of several equations that we call a meta-model. The invariance hypothesis is needed to ensure that the proportionality constant is the same for all assets. In this approach, the square root model is a knife-edged special case for which the invariance hypothesis is not needed for the proportionality constant to be the same for all assets. Kyle and Obizhaeva (2017*b*) illustrate this meta-model approach in the context of a concrete dynamic model of speculative trading with linear market impact.

The second *"dimensional analysis"* approach keeps the model of market impact as general as possible by beginning with dimensional analysis and then adding restrictions related to *leverage neutrality* and *market microstructure invariance*. It makes it possible to generalize the market impact formula by adding other explanatory variables such as execution horizon, minimum tick size, and minimum lot size. This approach is proposed by Kyle and Obizhaeva (2017*a*).

The meta-model approach and the dimensional analysis approach are complimentary to each other. When applying dimensional analysis, a common problem for all scientific fields is picking a correct, well-specified list of explanatory variables; indeed, the results often strongly depend on this initial choice. For example, we show in this paper that applying dimensional analysis incorrectly leads to empirically unrealistic predictions. This is why deriving predictions based on dimensional analysis is much less straightforward than it may seem to be ex post. In this sense, the meta-model approach has advantages that the dimensional analysis approach does not have. It imposes discipline on the set of explanatory variables to be included into the formula. It also generates specific predictions for the proportionality coefficient in the market impact function, such as equation (2). This coefficient turns out to be a specific non-linear function of the dollar cost of a bet and moment ratios of the shape of the distribution of bet sizes. Under the assumptions of market microstructure invariance, this coefficient is predicted to be the same or almost the same across all assets and markets.

At the same time, the dimensional analysis approach has flexibility that the meta-model approach lacks. If new asset characteristics are proposed as additional explanatory variables to be added to a universal market impact formula, dimensional analysis shows how this can be done in a natural manner by rotating the new parameters to make them dimensionless, scaling them to make them leverage neutral, and adding them as arguments to the function f in equation (2).

In this paper, we do not solve the market impact puzzle. Yet, by imposing more discipline on modeling and estimating the market impact function, we intend to make the market impact puzzle easier for researchers to solve in the future.

2 A Meta-Model Approach Based on Economics

Restrictions Based on Volume and Volatility Equations. A fully-specified economic model of market microstructure usually makes assumptions about assets' payoffs, the competitiveness of market makers and other traders, the distribution of private information across traders, and the trading rules which determine how prices are formed. Instead of developing such an equilibrium model, like Kyle and Obizhaeva (2017*b*), we identify some generic reduced-form properties that any fully specified model should have. These assumptions impose important restrictions on the universal market impact function.

Institutional investors execute bets or metaorders. For simplicity, suppose the percentage market impact G of executing a bet of size $|Q|$ is described by a power function

$$G = \alpha\,|Q|^{\beta}, \tag{3}$$

Here, market impact G is a dimensionless quantity, and bet size $|Q|$ has units of shares. The parameter α is a dimensional coefficient that may depend on some asset-specific characteristics such as volume and volatility, and β is the exponent of the market impact function. For example, execution of a buy bet of $Q = 5000$ shares may be expected to move prices up by $G = 10$ basis points. Bet size is drawn from a distribution symmetric about zero, with $\mathrm{E}\{Q\} = 0$. For buy bets, Q is positive; for sell bets, Q is negative. Let γ denote the expected number of bets (per day).

Trading volume depends on the size and number of bets, and returns volatility results from their market impact. Let V denote expected trading volume (shares per day), let σ^2 denote expected returns variance, and let P denote the stock price. We next introduce several intuitive economic restrictions and then discuss the implications that these restrictions impose on how the unknown parameter α relates to easily observable and calculated asset characteristics such as volume V, price P, and volatility σ.

Consider the reduced-form system of three simple equations which relate the size, number, and cost of bets to volume and volatility via the market impact that bets create:

$$
\begin{cases}
V = \gamma \, \mathrm{E}\{|Q|\}, & (4)\\
\sigma^2 = \gamma \, \mathrm{E}\{G^2\}, & (5)\\
\mathrm{E}\{G^2\} = \alpha^2 \, \mathrm{E}\{|Q|^{2\beta}\}. & (6)
\end{cases}
$$

The "volume equation" (4) defines expected trading volume as the product of the expected number and size of bets. The "volatility equation" (5) simply says that returns variance results from the market impact of independently distributed bets. The "market impact equation" (6), obtained by squaring the market impact function (3), measures the contribution to returns variance expected from one bet. In many finance models, prices follow martingales, at least approximately; therefore, we assume that the size, direction, and market impact of bets are approximately independently and identically distributed.[1]

These three equations place log-linear restrictions on five hard-to-measure quantities: γ, $\mathrm{E}\{|Q|\}$, $\mathrm{E}\{|Q|^{2\beta}\}$, $\mathrm{E}\{G^2\}$, and α. For example, α can be solved in terms of $\mathrm{E}\{|Q|\}$ and $\mathrm{E}\{|Q|^{2\beta}$ as

$$
\alpha = \sqrt{\frac{\sigma^2}{PV} \, \mathrm{E}\{|PQ|\}} \, \frac{1}{\sqrt{\mathrm{E}\{|Q|^{2\beta}\}}}, \tag{7}
$$

or in terms of γ and $\mathrm{E}\{|Q|^{2\beta}\}$ as

$$
\alpha = \frac{\sigma}{\sqrt{\gamma}} \frac{1}{\sqrt{\mathrm{E}\{|Q|^{2\beta}\}}}. \tag{8}
$$

Given equation(8), the system (4)–(6) implies the trading cost

$$
G = \frac{\sigma}{\sqrt{\gamma}} \frac{|Q|^{\beta}}{\sqrt{\mathrm{E}\{|Q|^{2\beta}\}}}. \tag{9}
$$

The last equation has simple intuition. A bet of size Q is converted into an unsigned z-score when $|Q|^{\beta}$ is scaled by its standard deviation $\sqrt{\mathrm{E}\{|Q|^{2\beta}\}}$. The scaled bet size is expressed in unsigned units of standard deviation. It measures how large an exponentiated bet $|Q|^{\beta}$ is relative to its average size, so that the standard deviation of the signed scaled bet size is equal to one. The equation implies that one bet is expected to move prices enough to create returns volatility of $\sigma/\sqrt{\gamma}$. The returns volatility per bet $\sigma/\sqrt{\gamma}$ is one way to quantify an asset's illiquidity. If γ is small, implying only few bets per day, then the asset is highly illiquid in the sense that each bet on average makes a large contribution to returns variance. For example, when daily returns volatility is 2 percent and about 100 bets are executed per day, as for some large U.S. stocks from the bottom of S&P 500 index, the measure of illiquidity is 20 basis points. When daily returns volatility is 3 percent and only one bet is usually executed per day, as for some very illiquid stocks, the measure of illiquidity is 3 percent, or 300 basis points.

It is useful to consider several special cases of the function (9) for different values of the ex-

[1]We can assume that $V = \zeta\gamma \, \mathrm{E}\{|Q|\}$ and $\sigma^2 = \phi\gamma \, \mathrm{E}\{G^2\}$. Here, ζ is an intermediation multiplier adjusting for non-bet trades executed by intermediaries; the parameter ϕ is a volatility multiplier adjusting for price changes resulting from non-bet events such as news announcements. Both multipliers are discussed by Kyle and Obizhaeva (2016) and Kyle, Obizhaeva and Kritzman (2016).

ponent β. If $\beta = 0$, we obtain a formula for the bid-ask spread, a fixed component of transaction costs that does not depend on how large a bet is:

$$G_0 := \alpha = \frac{\sigma}{\sqrt{\gamma}} = \sqrt{\frac{\sigma^2}{PV}\, P\, \mathrm{E}\{|Q|\}}. \tag{10}$$

This measure of transaction costs is typically most relevant for retail investors. If there are market frictions, such as monopolistic behavior by market makers, transaction taxes, or frictions which allow high frequency traders to profit at the expense of other traders, these frictions may show up as bid-ask spread costs. The next two measures of transaction costs, both of which depend on the size of bets, are usually of interest for institutional investors, who trade large quantities and for whom market impact costs largely dominate any bid-ask spread costs.

If $\beta = 1$, we obtain a formula for a linear market impact function, which is often seen in the theoretical market microstructure literature:

$$G_1 := \alpha\,|Q| = \frac{\sigma}{\sqrt{\gamma}}\,\frac{|Q|}{\sqrt{\mathrm{E}\{|Q|^2\}}} = \sqrt{\frac{\sigma^2}{PV}\, P\, \mathrm{E}\{|Q|\}}\,\frac{|Q|}{\sqrt{\mathrm{E}\{|Q|^2\}}}. \tag{11}$$

If $\beta = 1/2$, we obtain the well-known formula for the square root market impact function:

$$G_{1/2} := \alpha\,|Q|^{1/2} = \sigma\sqrt{\frac{|Q|}{V}}. \tag{12}$$

There is an important observation to be made about how easy these formulas are to implement empirically. In addition to depending on the easily observed variables volume V and volatility σ, these market impact formulas depend on *either* the number of bets γ *or* their average bet size $\mathrm{E}\{|Q|\}$ as well as its 2β-moment $\mathrm{E}\{|Q|^{2\beta}\}$. All of these variables are difficult to observe, and they are likely to vary significantly across assets.

The only exception is the square root model (12), in which the coefficient α depends solely on volume V and volatility σ because average bet size $\mathrm{E}\{|Q|\}$, its 2β-moment $\mathrm{E}\{|Q|^{2\beta}\}$, and γ all cancel out! Mathematically, this cancelation occurs because the two moments $\mathrm{E}\{|Q|\}$ and $\mathrm{E}\{|Q|^{2\beta}\}$ become the same when $\beta = 1/2$. In the square root case, knowledge of trading volume V is sufficient for deriving the parameter $\alpha = \sigma/\sqrt{V}$.

In the general case, one needs to know γ and $\mathrm{E}\{|Q|\}$, and this requires more information about the structure of order flow than just trading volume alone. Furthermore, knowledge of $\mathrm{E}\{|Q|^{2\beta}\}$ requires more information about the shape of bet size distributions. The intuition is simple. Prices move up and down as a result of buying and selling pressure measured by signed order flow. It is therefore important to know not only trading volume, but also its components—the number of bets and their average size—in order to interpret correctly the market impact and volatility equations.

Here is the problem. Suppose we fix some daily volume V and volatility σ. In principle, the same daily volume may result either from a few bets of large average size or many bets of small average size. For example, one million shares traded per day can represent either 100 trades of 10,000 shares or 10,000 trades of 100 shares. To keep daily volatility fixed, the implied impact per share must be ten times larger in the latter case because uncorrelated market impacts of a larger number of smaller bets would otherwise cancel one another out and lead to lower daily volatility. To illustrate, suppose that market impact is linear. If order flow consists of 100 trades of 10,000 shares, then each trade must have a market impact of 20 basis points ($0.0200/100^{1/2}$). If the order flow consists of 10,000 trades of 100 shares, then each trade has a market impact of 2 basis points ($0.0200/10,000^{1/2}$). In this latter case, linearity implies that a 10,000-share trade would have market impact of 200 basis points (2 basis points times 100). This is ten times more than the 20-basis-point market impact obtained in the first case.

In order to make empirical predictions about transaction costs, it is therefore necessary to have another restriction on the composition of the order flow. This restriction imposes a relationship that describes how the number of bets γ and average bet size $\mathrm{E}\{|Q|\}$ vary with volume V and volatility σ. Together with knowledge of moment ratios for bet size distributions, which are not affected by scale and defined below in equations (18), such a restriction would help pin down the market impact function. Without such a relationship, implementation of a market impact model, other than the square root model, would require estimating parameters describing bet size and the number of bets on a stock-by-stock basis. This is a daunting empirical challenge since bets are hard to observe.

Restrictions Implied by Transaction Cost Invariance. Market microstructure invariance solves the problem of determining bet size and number of bets by adding one more restriction to the equation system (4)–(6). The idea is to impose a restriction that links the number of bets γ and average bet size $\mathrm{E}\{|Q|\}$ to a new variable, which is expected to be approximately the same "invariant" parameter for all assets and therefore can be more easily estimated or calibrated.

There are two different ways of defining market microstructure invariance. We call them "transaction cost invariance" and "bet size invariance." These invariance hypotheses were first introduced by Kyle and Obizhaeva (2016). Let us consider transaction cost invariance first.

Transaction costs invariance hypothesizes that the ex ante expected dollar cost $\mathrm{E}\{|PQ|G\}$ of executing a bet, without conditioning on the size of a bet, satisfies the restriction

$$C = \mathrm{E}\{|PQ|G\}, \tag{13}$$

Taken literally, the invariance hypothesis implies that C is invariant across assets. Taken less strictly, it implies that C may vary somewhat across assets but not nearly as much as the number of trades γ or average trade size $\mathrm{E}\{|Q|\}$. For example, the value $C = \$2,000$ is approximately consistent with transaction cost estimates Kyle and Obizhaeva (2016) obtain from portfolio transitions. Interpreted strictly, invariance implies that the average cost of a bet is $2,000 for any market—equities, fixed income, 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 justifies our assumption about invariance of C.

Kyle and Obizhaeva (2017*b*) examine this idea using a specific dynamic infinite-horizon model with informed trading, noise trading, market making, endogenous production of information, and linear market impact. There is an asset with fundamental value following geometric Brownian motion. Risk-neutral traders arrive randomly. Their trades are intermediated by risk-neutral market makers. Each informed trader pays a fixed dollar cost for an informative private signal about the fundamental value. Each noise trader gets a similar "fake" signal. Noise traders are assumed to turn over an exogenously given fraction of the outstanding shares. The number of informed traders adjusts endogenously so that they expect to make zero profits, net of transaction costs and costs of obtaining private signals. The model assumes that the cost of acquiring information is the same across markets. Given this assumption, the model shows that the cost of information is the same or proportional to the market impact of trading on it. Thus, in equilibrium the average dollar market impact cost C of trading on each piece of information is the same.

Substituting the functional form of G from equation (3) into the invariance restriction (13) yields the equation $C = \alpha P\, \mathrm{E}\{|Q|^{1+\beta}\}$, and the system of three equations (4)–(6) becomes the following system of four equations:

$$
\begin{cases}
V = \gamma\, \mathrm{E}\{|Q|\}, & (14)\\
\sigma^2 = \gamma\, \mathrm{E}\{G^2\}, & (15)\\
\mathrm{E}\{G^2\} = \alpha^2\, \mathrm{E}\{|Q|^{2\beta}\}, & (16)\\
C = \alpha\, P\, \mathrm{E}\{|Q|^{1+\beta}\}. & (17)
\end{cases}
$$

This system of four equations has six unknowns: γ, α, $\mathrm{E}\{G^2\}$, $\mathrm{E}\{|Q|\}$, $\mathrm{E}\{|Q|^{1+\beta}\}$ and $\mathrm{E}\{|Q|^{2\beta}\}$. The last three unknowns are related to the shape of the distributions of bet size.

Let us introduce two dimensionless moment ratios m and m_β related to distributions of bet size $|Q|$:

$$
m := \frac{\mathrm{E}\{|Q|\}\sqrt{\mathrm{E}\{|Q|^{2\beta}\}}}{\mathrm{E}\{|Q|^{\beta+1}\}}, \qquad m_\beta := \frac{\left(\mathrm{E}\{|Q|\}\right)^{\beta+1}}{\mathrm{E}\{|Q|^{\beta+1}\}}. \tag{18}
$$

Since the exponents of $|Q|$ in the numerator and denominator are the same, the moment ratios m and m_β will be constant across assets as long as any two distributions of Q across assets differ only by a scalar. The moment ratios m and m_β are expected to be invariant constants which can be calibrated by pooling data across many assets. For example, Kyle and Obizhaeva (2016) show that distributions of portfolio transition orders in the U.S. stock market closely resemble log-normal distributions with a log-variance of approximately 2.50 and log-means varying across markets in a manner consistent with the hypothesis that C is constant across U.S. stocks. If the distributions are indeed log-normal with the same log-variance 2.50 and different log-means, then $m = \exp(2.50\,(\beta^2 - 2\beta)/2)$ and $m_\beta = \exp(-2.50\,\beta(\beta+1)/2)$ are reasonable empirical assumptions. If $\beta = 1/2$, then $m_\beta = m \approx 0.40$. If $\beta = 1$, then $m \approx 0.25$ and $m_\beta = m^2$.

We now have three easily observable characteristics P, V, and σ, which vary greatly across assets, and three characteristics C, m, and m_β, which are likely to be approximately constant across assets. In terms of these characteristics, the solution for $\mathrm{E}\{|Q|\}$, γ, and α is[2]

$$\mathrm{E}\{|Q|\} = (Cm)^{2/3}\,\frac{(\sigma PV)^{1/3}}{\sigma P}, \tag{19}$$

$$\gamma = (Cm)^{-2/3}\,(\sigma PV)^{2/3}, \tag{20}$$

$$\alpha = \frac{C^{(1-2\beta)/3}\,m_\beta}{m^{2(1+\beta)/3}}\,P^\beta \left(\frac{\sigma^2}{PV}\right)^{(1+\beta)/3}. \tag{21}$$

The exponents of 1/3 and 2/3 in equations (19) and (20) show how the size and number of bets in the order flow vary with volume and volatility given the transaction costs invariance assumption. A larger trading volume PV is 1/3 due to larger average bet size $\mathrm{E}\{|PQ|\}$ and 2/3 due to a larger number of bets γ. For example, holding price and volatility constant, 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 ($= 8^{2/3}$).

Thus, the invariance restriction implies that order flow has a very particular structure. This makes it possible to infer the unobservable structure of the order flow from observable dollar volume and volatility.

As mentioned earlier, there is something knife-edged about the way in which the square root model treats order imbalances. To see this, use equations (8), (18), and (20) to write market impact G as

$$G = \frac{m_\beta}{m}\,\gamma^{\beta-1/2}\,\sigma\left(\frac{|Q|}{V}\right)^\beta. \tag{22}$$

For the square root model $\beta = 1/2$, the exponent of γ is zero; therefore, all that is needed to model market impact and obtain the square root impact formula $G = \sigma\sqrt{(|Q|/V)}$ is information about volume V and volatility σ. For all other cases $\beta \neq 1/2$, the exponent of γ is non-zero; therefore, information about the number of bets γ (or average bet size) is needed to model impact in a manner that admits minimal economic restrictions. This is important because some studies find that market impact functions may have exponents closer to 0.60 than 1/2.

Using equation (20), the market impact (22) can be written as

$$G = \frac{m_\beta C^{(1-2\beta)/3}}{m^{2(1+\beta)/3}}\left(\frac{\sigma^2}{PV}\right)^{(\beta+1)/3} |PQ|^\beta. \tag{23}$$

The estimates of transaction costs for a bet Q depend only on the easily observable characteristics P, V, σ, and a proportionality constant which depends on C, m and m_β. This proportionality constant and exponent β can be estimated empirically. Equation (23) nests the square root model $\beta = 1/2$; in this special case, the constant term is equal to one even without the invariance assumption (since $m_\beta = m$ and the exponent of C is zero).

The assumptions that C, m and m_β are invariant are not necessary for derivation of the market impact formula (23), but they make it empirically operational since it becomes necessary to estimate only one constant which is the same for all assets. In this sense, the invariance

[2] We leave it as an exercise for the reader to derive the solution for the fourth parameter $\mathrm{E}\{G^2\}$.

assumption helps to solve the market impact puzzle by implying a universal market impact formula.

Let $1/L$ denote 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 $1/L = 0.0030$. Mathematically, the illiquidity measure $1/L$ is defined as the ratio of the expected dollar cost of executing a bet C to the expected dollar volume of the bet $\mathrm{E}\{|PQ|\}$,

$$\frac{1}{L} := \frac{C}{\mathrm{E}\{|PQ|\}} = \left(\frac{\sigma^2 C}{m^2 PV}\right)^{1/3}. \tag{24}$$

The last equation is obtained using equation (19). A more liquid market is associated with higher dollar volume PV and lower volatility σ.

The solution for $\mathrm{E}\{|Q|\}$ and γ in (19)–(20) can be expressed in a simple manner using L as

$$\mathrm{E}\{|PQ|\} = C\,L, \qquad \gamma = \frac{1}{m^2}\sigma^2 L^2. \tag{25}$$

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 $\mathrm{E}\{|PQ|\}$.

Using notation from Kyle and Obizhaeva (2017*a*), we can also write the market impact formula $G = \alpha\,|Q|^{\beta}$ as the product of two factors $1/L$ and $f(Z)$:

$$G = \frac{1}{L} f(Z). \tag{26}$$

Here, Z scales bet size by its mean and f is a specific invariant function of Z:

$$Z := \frac{Q}{\mathrm{E}\{|Q|\}} = \frac{PQ}{CL} \qquad \text{and} \qquad f(Z) = m_\beta |Z|^{\beta}. \tag{27}$$

The formula (26) is an important specification for transaction costs. It shows that percentage transaction costs G can be decomposed into the product of an asset-specific illiquidity index $1/L$ and an invariant function $f(\,)$ of scaled bet size $Z := Q/\mathrm{E}\{|Q|\}$, which measures how large a bet is relatively to an average bet. 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 $f(Z)$.[3]

To illustrate, assume linear market impact $f(Z) = m_\beta Z$, with $m = 0.25$ and $m_\beta = m^2 = 0.0625$. In a given market with $1/L = 0.0100$ or 100 basis points, 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 ($G = 0.000625Z$). In a different market, which is twice less illiquid with $1/L = 0.0050$ or 50 basis points, 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

[3] We have presented our derivation under the assumption of power market impact functions, but a similar decomposition (26) can be derived for any functional form g of asset-specific market impact functions. The more general case implies $f(Z) := g(Z)/\mathrm{E}\{|Z|\,g(Z)\}$, with $g(Z) = |Z|^{\beta}$ for the power function.

points. In both of these examples, a bet 10 times greater than average has a cost less than the average cost $1/L$. This occurs because the distribution of bet size has such enormous kurtosis that average transaction costs are dominated by gigantic bets even more than ten times greater than average bets (with $Z > 10$).[4]

Restrictions Implied by Bet Size Invariance. Instead of using transaction cost invariance, restrictions can also be generated by an alternative invariance hypothesis regarding the size of bets rather than their market impact. *Bet size invariance* hypothesizes that the dollar risk a bet transfers per unit of business time,

$$I := PQ\frac{\sigma}{\sqrt{\gamma}}, \tag{28}$$

has an invariant mean $\mathrm{E}\{|I|\}$ for all markets. The nominal size of a bet is PQ. If the expected arrival rate of bets sets the pace of business time, the returns standard deviation per unit of business time is $\sigma/\sqrt{\gamma}$. We can think of their product as the dollar risk transferred by a bet per unit of business time.

It can be shown that the transaction cost and bet size invariance hypotheses are closely related to each other. If we start with the definition $C = \mathrm{E}\{|PQ|\,G\}$, then substitute into it the market impact G from equation (9) and the moment ratio m from equation (18), we obtain the following relationship between the two hypotheses:

$$C = \frac{1}{m}\,\mathrm{E}\{|PQ|\}\,\frac{\sigma}{\sqrt{\gamma}} = \frac{1}{m}\,\mathrm{E}\{|I|\}. \tag{29}$$

The average dollar cost C is equal to the product of average bet size $\mathrm{E}\{|PQ|\}$ and average percentage impact $\sigma/\sqrt{\gamma}$ per bet, with the factor $1/m$ adjusting for differences in moments.

Thus, one could change the system (14)-(17) by replacing the cost invariance equation $C = \mathrm{E}\{|PQ|G\}$ with the bet size invariance equation $\mathrm{E}\{|I|\} = \mathrm{E}\{|PQ|\}\,(\sigma/\sqrt{\gamma})$ and obtain the same results, except with the invariant parameter C replaced by the invariant parameter $\mathrm{E}\{|I|\}/m$ in a manner consistent with equation (29).

If one assumes that not only the mean $\mathrm{E}\{|I|\}$ 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.

To summarize, the volume and volatility equations augmented with the hypotheses about invariance of average dollar costs and the shape of bet size distributions significantly constrain the possible specifications of market impact formulas and lead to operational formulas for transaction costs.

[4] The first example $1/L = 0.0100$ is consistent with equation (24) with the assumptions $C = \$2000$, volatility $\sigma =$ 2% per day, and dollar volume $PV = \$12,800,000$ per day. The second example $1/L = 0.050$ 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.

3 A Dimensional Analysis Approach with Leverage Neutrality

In physics, researchers obtain powerful results by using dimensional analysis to reduce the dimensionality of problems. Kyle and Obizhaeva (2017*a*) apply a similar approach to derive scaling laws for market microstructure.

In finance, dimensional analysis begins with identification of primary dimensions: asset "quantity" (units of shares and contracts), asset "value" or currency (units of dollars, pennies, pounds, and rubles), and "time" (units of weeks, days, hours, and seconds). All variables have dimensions which are products of powers of these primary dimensions. A variable is dimensionless if all of the exponents are zero. It is important for any proposed relationship among variables to be consistent with the three primary dimensions in which they are measured.

To further reduce degrees of freedom, Kyle and Obizhaeva (2017*a*) propose imposing an additional restriction to ensure that formulas are consistent with leverage neutrality, which is based on the following intuitive economic argument. Trading risky securities is costly. Exchanging cash-equivalent assets can be done at no cost since cash has no risk. The leverage neutrality hypothesis says that the economic costs of trading bundles of risky securities and a cash-equivalent asset are the same regardless of any positive or negative amount of cash-equivalent assets included in a bundle. Changes in leverage, margin requirements, or repo haircuts do not alter fundamental relationships among variables. For example, suppose an amount of cash equal to aP is added to each share of stock. This decrease in leverage raises the stock price to $(1+a)P$, lowers returns volatility to $\sigma/(1+a)$, and reduces percentage bid-ask spread to $(1/(1+a))\, s/P$, but it does not change V, $\mathrm{E}\{|Q|\}$, or γ.[5]

Suppose there exist primary asset characteristics that constitute *a basis*, so that market microstructure variables can be expressed as some functions of these characteristics. It is intuitive to assume this basis includes price, volume, and returns volatility $\{V, P, \sigma^2\}$; it is relatively easy to observe or calculate all three variables. Following Pohl et al. (2017), we summarize information about dimensions and units in matrix form. In the header, we have dimensions: M (value, money, currency), S (quantity), T (time), with L (leverage) in the last column. In the first three rows, we have variables P, V, and σ with their dimensions. For example, stock price P has units of dollars per share, and changes inversely with leverage. Trading volume V has units of shares per day and does not change with leverage. Returns variance σ^2 has units per-day and changes twice fast with leverage. The remaining rows contain other potential variables of interest, such as average bet size $\mathrm{E}\{|Q|\}$, number of bets γ, percentage spread s/P, and market impact G.

[5] For example, consider the costs of trading a futures contract on an index. If the margin requirement for a futures contract with price P is 10%, then the trader might consider the value of the contract to be $0.10\,P$, consistent with the assumption $a = -0.90$. The futures contract is priced to take into account financing of 90% of its value. In comparison with the underlying contract itself (with margin of 100%), leverage neutrality implies that the trading volume V, size of bets $\mathrm{E}\{|Q|\}$, number of bets γ, and dollar trading costs are not affected by this leverage implied by margin requirements, but the percentage volatility of the margin account is 10 times greater than the volatility of the unlevered contract, and trading costs (including percentage bid-ask spread) measured as a fraction of the value of the margin account are 10 times greater as well.

	M	S	T	L
price P	1	-1	0	-1
volume V	0	1	-1	0
variance σ^2	0	0	-1	2
average bet $\mathrm{E}\{\lvert Q\rvert\}$	0	1	0	0
number of bets γ	0	0	-1	0
percentage spread s/P	0	0	0	1
market impact G	0	0	0	1

Suppose we would now like to express average bet size $\mathrm{E}\{|PQ|\}$, the number of bets γ, and percentage bid-ask spread s/P as functions of the three characteristics price P, volume V, and volatility σ. Let the functions be denoted f_q, f_γ, and f_s, respectively. Dimensional consistency (the Buckingham π Theorem) leads immediately to the functional forms

$$\mathrm{E}\{|PQ|\} = f_q(P,V,\sigma) \sim \frac{PV}{\sigma^2} \qquad \text{and} \qquad \gamma = f_\gamma(P,V,\sigma) \sim \sigma^2 \qquad \text{and} \qquad \frac{s}{P} = f_s(P,V,\sigma) \sim 1. \tag{30}$$

Because there are three dimensions to match consistently and only three characteristics (P, V, σ), there are not enough degrees of freedom to match both the three restrictions of dimensional consistency as well as leverage neutrality.

Not surprisingly, none of these three predictions (30) is consistent with leverage neutrality. Furthermore, it is implausible to imagine that these predictions might be consistent with the data. For example, it is unreasonable to expect that the number of bets γ is proportional to returns variance σ^2. Indeed, volatility is relatively stable across stocks, ranging from about 1.5 to 3 percent per day for U.S. stocks, whereas the number of bets γ is likely to change by orders of magnitude across stocks. The percentage spread is known to vary greatly across markets, whereas it is predicted to be constant in equation (30). Kyle, Obizhaeva and Tuzun (2010), Andersen et al. (2015), Kyle and Obizhaeva (2016), and Benzaquen, Donier and Bouchaud (2016) study empirical relationships among these variables. Kyle et al. (2010) and Bae et al. (2014) analyze similar relationships for the number of news articles and the number of switching points.

Suppose next that market impact G is a function of the three primary variables P, V, σ, and the size of the bet Q. Since a fourth variable Q has been added to the mix, one obtains a unique function that satisfies both dimensional consistency and leverage neutrality, up to some undetermined proportionality constant. This unique result

$$G = f_g(P,V,\sigma;Q) \sim \sigma\sqrt{\frac{|Q|}{V}} \tag{31}$$

is the square root model of market impact; this point is emphasized by Pohl et al. (2017). This prediction is consistent with leverage neutrality because percentage market impact G and returns volatility σ both change linearly with leverage. As we discussed, this prediction seems to be broadly consistent with the data. The restriction of the basis in dimensional analysis to $\{V,P,\sigma^2\}$ leads to the square root model, which itself is consistent with leverage neutrality. Why may this restricted basis work for the market impact formula but not for other formulas? Is it just a coincidence or does there exist some deeper explanation requiring further investigation?

This discussion illustrates that application of dimensional analysis is not simple. Results depend crucially on the initial decision about what variables are included in the list of the function's arguments. Dimensional analysis may lead to incorrect results if one starts with an incorrect set of arguments or omits some variables.

What is a correct basis in market microstructure? Which variables, if any, must one add to the basis $\{V, P, \sigma^2\}$? First, these variables could be variables that are difficult to observe or calculate, such as moments of bet sizes and number of bets. Second, these variables could be some easily observable variables such as number of trades, moments of trade sizes, market float, market capitalization, minimum tick size, or minimum lot size. Third, these could be variables that are expected to be relatively constant across different markets so that they show up in a proportionality constant such as average dollar cost C and the moment ratios m and m_β. To obtain operational formulas, it is preferable for the additional variables to be easily observed or relatively constant across assets, but identification of correct basis variables is ultimately an empirical question.

Researchers often guess basis variables ether by looking at results of experiments or analyzing insights from theoretical modeling. In finance, experiments are rare. The next table contains possible candidates that show up in our meta-model in the previous section.

	M	S	T	L
price P	1	-1	0	-1
volume V	0	1	-1	0
variance σ^2	0	0	-1	2
X	μ	s	t	l
bet cost C	1	0	0	0
moment ratio m	0	0	0	0
bet size moment $\mathrm{E}\{\lvert Q\rvert^{\beta+1}\}$	0	$\beta+1$	0	0
number of bets γ	0	0	-1	0

The first three rows contain price P, volume V, and variance σ^2. The next row contains a generic variable X with arbitrary dimensions μ, s, t, l. Since the three variables P, V, σ^2 span all three primary dimensions, we multiply X by powers of the three variables P, V, σ^2 to rotate X and make it dimensionless. We then exponentiate the rotated dimensionless variable so that it scales linearly with leverage:

$$\left(\frac{X\,\sigma^{2(\mu+s+t)}}{P^{\mu}\,V^{\mu+s}}\right)^{1/(l+\mu+2(\mu+s+t))} \tag{32}$$

This procedure is always possible unless the rotated variable has no leverage, in which case the denominator of the exponent $l+\mu+2(\mu+s+t)$ is zero. The variable X together with price, volume, and returns variance creates a new basis $\{V, P, \sigma^2, X\}$.

For example, if X is the number of bets per day γ, then $\mu = 0$, $s = 0$, $t = -1$, and $l = 0$, and the scaled variable is $(\sigma^2/\gamma)^{1/2}$. If X is average bet size $\mathrm{E}\{|Q|\}$, then $\mu = 0$, $s = 1$, $t = 0$, and $l = 0$, and the scaled variable is $(\mathrm{E}\{|Q|\}\sigma^2/V)^{-1/2}$. Both of these scaled variables are difficult to observe, and they most likely vary across markets.

If X is the average dollar cost of executing a bet C, then $\mu = 1$, $s = 0$, $t = 0$, and $l = 0$, and the scaled variable is

$$\left(\frac{C\sigma^2}{PV}\right)^{1/3}. \tag{33}$$

This variable is proportional to the illiquidity measure $1/L$, defined in equation (24), with a dimensionless proportionality coefficient $m^{-2/3}$. With the variable C added to price, volume, and volatility, the new basis $\{V, P, \sigma^2, C\}$ can be used to span any dimension of quantity, value, time and to match any amount of leverage. This is the basis considered by Kyle and Obizhaeva (2017*a*).

As we mentioned earlier, there are good economic reasons to believe that even though C is hard to observe, it is likely to be relatively constant across markets. Assuming the basis $\{V, P, \sigma^2, C\}$ and invariance of average bet cost C, we can apply dimensional analysis, leverage neutrality, and invariance to obtain the market impact formula

$$G = g(\sigma, P, V, C; Q) \quad \sim \quad \left(\frac{\sigma^2}{PV}\right)^{1/3} f\left(\left(\frac{\sigma^2}{PV}\right)^{1/3} |PQ|\right), \tag{34}$$

as shown in more detail by Kyle and Obizhaeva (2017*a*). If additional parameters $Y_1, \ldots, Y_n$ have already been scaled to be dimensionless and leverage neutral, this approach generalizes market impact equations (26) and (34) to obtain

$$G = \frac{1}{L} f(Z, Y_1, \ldots, Y_n), \qquad \text{where} \qquad \frac{1}{L} := \frac{\sigma^2 C}{m^2 PV} \qquad \text{and} \qquad Z := \frac{PQ}{CL}. \tag{35}$$

Determining the specific extra variables needed $(Y_1, \ldots, Y_n)$ is ultimately an empirical question.

4 Conclusions and Future Research

Both general approaches in this paper lead to a market impact function which is the product of an asset-specific illiquidity measure $1/L$ and an invariant function of scaled bet size $f(Z)$, consistent with equations (2), (23) and (34).

Under both approaches, the square root model is a special knife-edged case of the general market impact formula. It is the only case for which the market impact function, including proportionality constants, depends only on volume and volatility and not on characteristics of the order flow like the number of bets, their average size, and some of its moments. In other cases, this additional information becomes important.

The three variables P, V and σ are not sufficient for obtaining reasonable predictions for the size and number of bets, the bid-ask spread, and other microstructure characteristics describing a particular market. This problem occurs because any product of powers of the three variables stock price P, volume V, and volatility σ that satisfies the three necessary dimensionality constraints will not necessarily satisfy the constraint of leverage neutrality. To satisfy leverage neutrality, another market characteristic must be added to the basis $\{V, P, \sigma^2\}$. Mathematically, it does not make a difference whether the added characteristic is number of bets γ, average bet size $\mathrm{E}\{|Q|\}$, or average bet cost C because the meta-model reveals that all of the three variables show up in theoretical restrictions. Our preference for C over the other two variables is based on the invariance hypothesis that cost C, together with moment ratios m and m_β, may be approximately constant across assets and therefore easier to calibrate. We could also include another invariant $E\{|I|\}$.

In both approaches, there are still many open questions. Our meta-model approach implicitly relies on the assumption that information generates discrete trades of large sizes so that one can identify bets and their market impact in the data. Extracting bets from the order flow may be empirically difficult when asset managers collect information continuously in many tiny pieces and use order shredding algorithms to trade continuously, as modeled by Kyle, Obizhaeva and Wang (2017). Perhaps a different meta-model approach is needed to model smooth trading. Another approach may also be needed to model order flow in highly fragmented markets, where bets may be split and sent for execution to numerous trading venues.

The dimensional analysis approach leaves many unanswered questions for further research as well. What is the set of correct arguments? Can some variables such as volume, price, and volatility be of the first-order importance while other variables such as minimum tick size or minimum lot size be less important? Is there some sense in which the number of bets, the size of bets, and the cost of bets are "more fundamental" than other variables like number of trades, average trade size, or number of news articles?

There are still many open questions related to market microstructure invariance. Are the variables C, m, and m_β approximately constant across markets, countries, and time periods? If so, what are their values? Or if not, alternatively, can one identify a set of regimes in which these variables are relatively constant? Are there other similar variables that can be almost invariant?

Finally, is there a theory based on financial economics which leads to a square root model of market impact? For example, assuming that expected execution horizon T increases linearly with scaled bet size $|Z|^{1/2}$ for each asset, Kyle and Obizhaeva (2017*a*) suggest that an empirically misspecified square root model may be obtained if order execution time T is left out of a linear model, and the correct specification is a linear function of scaled bet size Z and inverse of scaled execution horizon B:

$$G \sim \frac{1}{L}\,\frac{|Z|}{B}, \qquad \text{where} \qquad \frac{1}{L} = \left(\frac{\sigma^2 C}{m^2 P V}\right)^{1/3}, \qquad Z := \frac{PQ}{CL}, \qquad B \sim T\,\sigma^2\,L^2. \tag{36}$$

This market impact function is linear in execution rate Z/B, where Z is the bet size scaled by the average bet size and B is the number of other bets which are expected to arrive while the bet is being executed, as derived using the dimensional analysis approach. For example, if a bet is 4 times bigger than the average bet $E\{|Q|\}$ and it is executed over a twice longer horizon than the average bet, then its market impact may be twice as large because its execution rate is twice as fast.

Is it possible for our approach to generate quantitative predictions about the dynamic properties of market impact? For example, market impact may be decomposed into permanent and temporary components. Both components represent market impact costs to the trader. The permanent component affects prices for a long time while the temporary component goes away over time. These components are difficult to identify empirically and difficult to model theoretically, as can be seen in the smooth trading model of Kyle, Obizhaeva and Wang (2017).

There are many open questions about the puzzle of transaction costs functions, which we consider to be one of the most interesting outstanding puzzles in finance.

References

Almgren, Robert, Chee Thum, Emmanuel Hauptmann, and Hong Li. 2005. “Direct Estimation of Equity Market Impact.” *Risk*, 18(7): 58–62.

Andersen, Torben G., Oleg Bondarenko, Albert S. Kyle, and Anna A. Obizhaeva. 2015. “Intraday Trading Invariance in the E-Mini S&P 500 Futures Market.” Working Paper. Available at http://dx.doi.org/10.2139/ssrn.2693810.

Bae, Kyounghun, Albert S. Kyle, Eun Jung Lee, and Anna A. Obizhaeva. 2014. “Invariance of Buy-Sell Switching Points.” Working Paper. Available at http://dx.doi.org/10.2139/ssrn.2730770.

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.

Gabaix, Xavier, Parameswaran Gopikrishnan, Vasiliki Plerou, and Eugene H. Stanley. 2006. “Institutional Investors and Stock Market Volatility.” *The Quarterly Journal of Economics*, 121: 461–504.

Grinold, Richard C., and Ronald N. Kahn. 1995. *Active Portfolio Management.* Probus Publishing.

Kyle, Albert S. 1985. “Continuous Auctions and Insider Trading.” *Econometrica*, 53(6): 1315–1335.

Kyle, Albert S., and Anna A. Obizhaeva. 2016. “Market Microstructure Invariance: Empirical Hypotheses.” *Econometrica*, 84(4): 1345–1404.

Kyle, Albert S., and Anna A. Obizhaeva. 2017*a*. “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. 2017*b*. “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., Anna A. Obizhaeva, and Mark Kritzman. 2016. “A Practitioner’s Guide to Market Microstructure Invariance.” *Journal of Portfolio Management*, 43(1): 43–53.

Kyle, Albert S., Anna A. Obizhaeva, and Tugkan Tuzun. 2010. “Microstructure Invariance in U.S. Stock Market Trades.” Working Paper. Available at http://dx.doi.org/10.17016/FEDS.2016.034.

Kyle, Albert S., Anna A. Obizhaeva, and Yajun Wang. 2017. “Smooth Trading with Overconfidence and Market Power.” *Review of Economic Studies*, Posted March 8: http://www.restud.com/paper/smooth–trading–with–overconfidence–and–market–power/.

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.

Loeb, Thomas F. 1983. "Trading Cost: The Critical Link Between Investment Information and Results." *Financial Analysts Journal*, 39(3): 39–44.

Obizhaeva, Anna A. 2012. "Liquidity Estimates and Selection Bias." Working Paper, New Economic School.

Pohl, Mathias, Alexander Ristig, Walter Schachermayer, and Ludovic Tangpi. 2017. "The Amazing Power of Dimensional Analysis: Quantifying Market Impact." Working Paper. Available at https://arxiv.org/abs/1702.05434.