Theoretical analysis of the bid-ask bounce and Related Phenomena

Lerner, Peter

Syracuse University (Ret.) and LIM College, NYC.

December 2010

RESEARCH ARTICLE

AESTIMATIO, THE IEB INTERNATIONAL JOURNAL OF FINANCE, 2010. 1: 144-163

Theoretical Analysis of the Bid-Ask Bounce

Lerner, Peter

▸ SUBMITTED : APRIL 2010
▸ ACCEPTED : NOVEMBER 2010

I provide a theoretical model for two empirical phenomena observed in the NYSE and Nasdaq markets. First is the bid-ask bounce recently studied by Heston, Korajczuk and Sadka (HKS, 2008) for high-frequency data. Second is a temporary liquidity squeeze observed by Madureira and Underwood (2008) in the event studies. The model I invoke to explain empirical observations of those two groups of authors, is based on Easley, Kiefer, O'Hara and Paperman (EKHP, 1996) equations for informed trading. The estimation was performed by maximizing correlations between MCMC-generated paths and empirical time series, which also maximizes the entropy.

My modeling rejects the rational expectation paradigm on a short-to-medium (15 min. to 2 days) time scale. I conclude that, given statistical uncertainty, roughly half of the bid-ask spread can be attributed to the arrival of new economic information and the other half to microstructure friction.

Market microstructure, EMH (Efficient Market Hypothesis), Nasdaq, High frequency finance, Autocorrelation of returns.

G14, G12, G19.

Lerner, P.B., ✉ *Syracuse University (Ret.) and LIM College, NYC. Ithaca, NY 14850, rdml134@earthlink.net*

Normally, one would expect that a "buy" order gets executed close to an ask price and a "sell" order near the bid price. If this were the case, the bid-ask spread would stay almost constant for quite some time until the next macroeconomic event. However, buyers frequently trade near a bid price, while sellers might sell near the offer. Because the orders get lumped together in reporting, it appears as though the market price oscillates between bid and ask bounds. This phenomenon is called bid-ask bounce (Harris, 2002). Furthermore, when news is released, market prices find a new level, sometimes monotonously but sometimes in a see-saw pattern (Plott and Smith, 2008, Chapter 1). Do market participants re-learn something they forgot in the course of the previous few minutes of trading? A related question of why market returns are autocorrelated even for very liquid markets has been extensively studied both theoretically and empirically (see Hasbrouck, 2007 for an extensive review). Another problem is the reaction of a bid-ask spread to a market event. Supply and demand of liquidity go to the core of economic foundations of financial economics. However, the theoretical description of the bid-ask bounce and event response remains sketchy and is discussed mostly in heuristic terms. The goal of this paper is to develop a stylized quantitative theory of the bid-ask bounce and autocorrelation that can, at least in principle, be calibrated.

My proposed method of estimation of the pricing model is close to the version of Bayesian inference long used in signal processing (Candy, 2009). This method is based on the fact (Granger and Lin, 1994) that the correlation coefficient can be interpreted in terms of entropy or Kullback-Leibler distance between distributions (Lawler, 2006, Hong, 2006). The structure of the paper is as follows. In Section 2 I remind the reader of the Easley, Kiefer, O'Hara and Paperman (EKHP, 1996) version of the Glosten-Milgrom model (1985) of informed trading. In the third section, I outline the empirical results of Heston, Korajczuk and Sadka (HKS, 2008). In the fourth section, I provide an outline of the results from signal processing theory, which were used for estimation in my paper. In Section 5, I provide the estimation of my model using empirical results of the HKS collaboration. Sections 3-5 represent an expanded and corrected version of the model provided in my book (Lerner, 2009). In Section 6, the reaction of the bid-ask spread to market events as described by Madureira and Underwood (2008) is explained in terms of the three-agent model. In Section 7, I discuss the limits of applicability for the proposed model.

Glosten and Milgrom (1985) formulated their model of Bayesian updates independently and probably slightly ahead of Kyle (1985). However, their approach is economic (game-theoretic) and does involve explicit dynamic equations. It was modified into potentially solvable form by Easley, Kiefer, O'Hara and Paperman, further quoted as EKHP (1996) and I shall restrict myself to that approach.

Below, I formulate and derive the Easley, Kiefer, O'Hara and Paperman (EKHP, 1996) modification of the Glosten-Milgrom model from the Bayes theorem. These equations can be derived in a purely algebraic fashion. Price evolution depends on update of beliefs in subsequent acts of trading.

Proposition 1.1 In a two-period, three-agent, one-asset market, the bid and ask prices a_1, b_1 obey the following system of equations:

$$a_1 = E(V_1|0) + \frac{\mu P_{g,0}}{\varepsilon + \mu P_{g,0}} (\overline{V}_1 - E(V_1|0))$$
$$b_1 = E(V_1|0) + \frac{\mu P_{b,0}}{\varepsilon + \mu P_{b,0}} (\underline{V}_1 - E(V_1|0)) \tag{1}$$

In Equation (1), P_b, P_g are the respective probabilities of bad and good news, and ε and μ are the rates of arrival of liquidity and insider traders on the market. $\overline{V}_1$ and $\underline{V}_1$ are the prices of an asset contingent on good and bad news, and $E(V_1|0)$ is an expected asset price given all the information preceding t=1. Equation (1) determines, in principle, dynamics of the bids and offers, if the processes for updates of prices and probabilities are known. From (1), for the period t+1, we get

$$a_{t+1} = E(V_{t+1}|t) + \frac{\mu P_{g,t}}{\varepsilon + \mu P_{g,t}} (\overline{V}_{t+1} - E(V_{t+1}|t))$$
$$b_{t+1} = E(V_{t+1}|t) + \frac{\mu P_{b,t}}{\varepsilon + \mu P_{b,t}} (\underline{V}_{t+1} - E(V_{t+1}|t)) \tag{2}$$

In this model of sequential trades, individuals trade a single risky asset, whose true value depends on the "good," "bad," and "neutral" states of nature. At the end of the trading period, the information about the true value of an asset is fully incorporated into prices. Equations (1) and (2) are derived in Appendix A.

If we substitute past ask and bid prices at time t into Equations (2), we obtain:

$$a_{t+1} = V_t \left(1 - \frac{\mu P_g}{\varepsilon + \mu P_g}\right)^2 + \left(\frac{2\mu P_g}{\varepsilon + \mu P_g} - \left(\frac{\mu P_g}{\varepsilon + \mu P_g}\right)^2\right)\overline{V}_t + \frac{\mu P_g}{\varepsilon + \mu P_g} \; u_{t+1}$$
$$b_{t+1} = V_t \left(1 - \frac{\mu P_b}{\varepsilon + \mu P_b}\right)^2 + \left(\frac{2\mu P_b}{\varepsilon + \mu P_b} - \left(\frac{\mu P_b}{\varepsilon + \mu P_b}\right)^2\right)\underline{V}_t + \frac{\mu P_b}{\varepsilon + \mu P_b} \; v_{t+1} \tag{3}$$

where $$u_{t+1}=\bar{V}_{t+1}-\bar{V}_t,$$
$$v_{t+1}=\underline{V}_{t+1}-\underline{V}_t, \tag{4}$$

are the asset price innovations at time $t+1$. From now on we consider the frequency of good and bad events as exogenously given constants. If we define new constants, according to the rule:

$$\eta=\frac{\mu P_g}{\varepsilon+\mu P_g}, \quad \tilde{\eta}=\frac{\mu P_b}{\varepsilon+\mu P_b} \tag{5}$$

the resulting equations obtain the form:

$$a_{t+1}=E(V_{t+1}|t)\,(1-\eta)^2/2+(2\eta-\eta^2)\bar{V}_{t+1}+\eta u_{t+1}$$
$$b_{t+1}=E(V_{t+1}|t)\,(1-\eta)^2/2+(2\tilde{\eta}-\tilde{\eta}^2)\underline{V}_t+\tilde{\eta}u_{t+1} \tag{6}$$

The EKHP equations in the form (5-6) will be used to demonstrate the bid-ask bounce and the event reaction in Sections below.

Here, I compare the model of Equations (6) with the empirical results of Heston, Korajczuk and Sadka (2008), further quoted as HKS. One advantage of their empirical methodology is that they analyze four years of returns of all NYSE-listed stocks in 16,261 half-hour intervals. HKS start their sample in January 2001 so as to coincide with the end of decimalization and end it in December 2005. The following model of excess returns was analyzed:

$$r_{i,t}=\alpha_t+\gamma_{t,k}\,r_{i,t-k}+u_{it} \tag{7}$$

where r_i is the return of stock i and t's are half-hour intervals throughout the period. In Equation (7), k is a lag index, $1\leq k\leq 65$, indicating a position of lag within a trading week.

In the opinion of the authors of HKS, possible alternative hypotheses for the serial correlations in the bid-ask spread, in order are: a) non-uniform release of financial information throughout the day, b) scheduling of the index funds trades at the end of the day to minimize tracking error, c) less liquid stocks and more liquid stocks are traded in different parts of the day, d) price pattern follows a pattern in volume,

e) intraday periodicity in volatility, f) fluctuations in imbalance, g) the difference between transaction and bid-ask prices and h) pre-decimalization results.

The authors rule out explanation a) by eliminating the market risk. The oscillations do not disappear. Stocks that are members of an index demonstrate a pattern with much smaller amplitude, which rules out explanation b). HKS measured size and transaction costs and established that if there were preferences for high- and low-liquidity stocks in different parts of the day, they would have been large enough to stimulate cross-trading. This cross-trading would eliminate the pattern.

HKS made a cross-sectional regression of volume on volume:

$$v_{it}=a_{tk}+g_{tk}\,v_{i,t-k}+u_{it} \tag{8}$$

where v is the volume, i is the stock number, t is a moment of time and k is a lag. Serial correlations in volume were not statistically significant to explain the price autocorrelation. The authors did not find periodicity in volume except on a daily frequency basis.

To eliminate the influence of the difference between the bid and ask, and transaction price, respectively, the authors computed returns with a) bid prices only, b) ask prices only and c) mid-price quotes. The results were similar.

Finally, HKS noticed that intraday periodicity increased after decimalization. To mitigate the results of shrinking quotations (1/8 for the period 1993-1997, 1/16 for 1997-2000 and decimals for 2001-2005), the authors introduced portfolio returns, from which the top and bottom 1% results were removed. Again, the autocorrelation results were similar for "raw" and truncated (winsorized) portfolios. The authors of HKS conclude that the above due diligence statistics "document pronounced intraday reversals due to bid/ask bounce."

A standard application of entropy in financial time series is a description of the affinity of distributions (Y. Hong, 2006). This application is based on the notion of Kullback-Leibler Distance (KLD, see e.g. Lawler, 2006) between two alternative distributions: $f_0(x,y)$, which we consider baseline and $f_1(x,y)$:

$$I_{01}=\mathrm{E}\left[\log\left(\frac{f_1(x,y)}{f_0(x,y)}\right)\right]=\int\log\left(\frac{f_1(x,y)}{f_0(x,y)}\right)f_1(x,y)\,dxdy \tag{9}$$

The intuitive meaning of the KLD is the information we obtain if, instead of the expected $f_0(x,y)$, the observed distribution is $f_1(x,y)$. Granger and Lin (1994) proposed a normalized entropy measure:

$$e_{01}^2 = 1 - \exp(-2I_{01}) \quad (10)$$

Normalized entropy (10) has obvious properties resulting from the properties of the KLD (e.g. Hong, 2006):

(a) $e_{01} = 0$ if and only if $f_0(x,y) = f_1(x,y)$;
(b) $e_{01} = 1$ if and only if y is functionally dependent on x;
(c) e_{01} is invariant under transformation $x' = h_1(x)$, $y' = h_2(y)$ where h_1 and h_2 are smooth monotonic functions;
(d) If $f_1(x,y,\rho)$ is Gaussian and $f_0 = f_1(x,y,\rho=0)$, then $e_{01} = |\rho|$.

My optimization procedure in Section 5 is based on property (d). Namely, if one has a random sample $\{\mathbf{X}_i\}$ with parameters θ_i then one can use normalized entropy as a distribution function. Optimizing this distribution, according to Granger and Lin (1994), will be equivalent to maximizing entropy.

This method has a computational advantage because rather than extracting distributions from the time series, then calculating the integral (8) and, finally, maximizing I_{01}, one can compute the correlation between the empirical sample and the simulated sample and then weigh observations according to $|\rho|$ in a single swoop.

I will not speculate on the economic reasons for the bid-ask bounce, which are usually explained by the hard-to-define concept of "immediacy" (O'Hara, 1995). Yet, from the modeling perspective, as in the original model of Section 2, it suffices that the agents are segmented into knowledgeable insiders, who can time the favorable moment for an execution, and liquidity traders who mostly act as price-takers. Bid-ask bounce can be simulated on the basis of the extended EKHP equations (A.1), which we derived in Appendix A. Because, in the absence of informed traders, the coefficient for $E[V_{t+1}|t]$ is equal to 1 and the coefficients for $\underline{V}_{t+1}$, $\overline{V}_{t+1}$ are equal to zero, one can crudely identify the first term with the impact of liquidity and the second term with the impact of informed trading. Note that, in the absence of the informed traders EKHP equations predict empirical martingale evolution of prices in accord with naive intuition. Generally speaking, the rule under which market agents predict the price can be arbitrary, but we can use the efficient market convention that the best estimator for the next transaction's price is the current (mid-) price:

$$E[V_{t+1}|t]=V_t=\frac{(a_t+b_t)}{2} \tag{11}$$

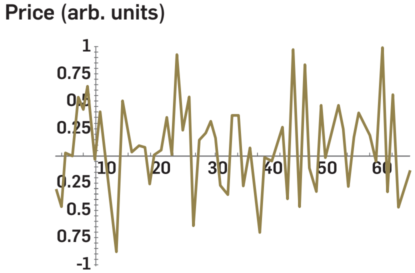
where $u_{t+1},v_{t+1}\sim N(0,\sigma_1^2)$, see Appendix A. As far as Equations (6) and (11) are concerned, we still do not know the price process that determines the upper and lower bounds of the future price. On a relatively short time scale, we can assume that the upper and lower bounds of asset price stay approximately constant, changing randomly in small increments in sync with positive and negative events:

$$\begin{gathered}\overline{V}_t=\overline{V}_0+\sqrt{\upsilon}\ max[\varepsilon_t,0]\\ \underline{V}_t=\underline{V}_0+min[\varepsilon_t,0]\end{gathered} \tag{12}$$

where $\varepsilon t\sim N(0,\sigma_2^2)$, υ is an adjustable parameter, indicating asymmetry of reaction to positive and negative news. A typical random path of the price evolution is shown in Figure 1. I must caution that at a large number of lags our model becomes inaccurate because simple approximation for the asset prices (12) fails. However, at a large number of daily lags, one may expect that all autocorrelations die out or overlap with the price movements from new information.

I do not know the exact solution of the model provided by Equations (6) and (11-12). Hence, any educated guess of the MLE is impossible. To estimate model parameters for the empirical situation of HKS, I use the method of Bayesian inference (Tsay, 2002), a variation of which I previously implemented for pricing of the EU pollution quotas (Lerner, 2008). Imagine that Markov Chain Monte Carlo (MCMC) simulations generate an empirical sample $X(\theta)$ for a distribution of model parameters θ. For a given value of parameter vector θ_0, I can generate a set of the model paths similar to Figure 1.

Ask prices from the model of Equations (6) and (11) for an arbitrary (not estimated) value of parameter vector θ_0 for 65 days



Units on the vertical axis are arbitrary as correlation in condition 3 (d) does not depend on the parameter's scale.

If one postulates the criteria for the likelihood of the model path and the empirical trajectory of the asset, in this case, a correlation coefficient, one may form an expression:

$$\rho(\theta_0|X(\theta)) = \frac{Cov[X(\theta_0), X]}{\sqrt{Var[X(\theta_0)]}\,\sqrt{Var[X]}} = \frac{\hat{E}[X(\theta_0) - \hat{E}[X(\theta_0)], X - E[X]]}{\sqrt{\hat{E}[\left(X(\theta_0) - \hat{E}[X(\theta_0)]\right)^2]}\,\sqrt{E\left[(X - E[X])^2\right]}} \tag{13}$$

where $\mathbf{X}$ is an empirical sample, expectation sign with a hat means an average over computer-generated paths, while the expectation sign without hat means the mean of the empirical sample. As is usual with Bayesian methods, we must select a prior $P(\theta)$, which can be any reasonable parametric distribution. In this case, I choose a gamma distribution. We can estimate a parameter vector from the Bayes formula:

$$\hat{\theta} = \underset{\theta_0 \in \Theta}{arg\ max} \int |\rho(\theta_0|\theta| P(\theta) d\theta \tag{14}$$

The logic behind Equation (14) is that we estimate a true vector of parameters as maximizing the posterior distribution of the correlation coefficient between the empirical sample and the model-generated paths. In actuality, I do not solve Equation (14), but estimate a vector of parameters by the mode of the posterior distribution of the vector of parameters:

$$\hat{\theta} \approx E_{\rho,B}[\theta P(\theta)] = \frac{\int \theta |\rho\ (\theta_0|\theta)| P(\theta) d\theta}{\iint |\rho(\theta_0|\theta)| P(\theta) d\theta} \tag{15}$$

where index B in the expectation sign means "Bayesian", i.e. the expectation according to prior distribution. Here, I am not concerned with the exact shape of the posterior distribution because I lack the HKS data to compare.

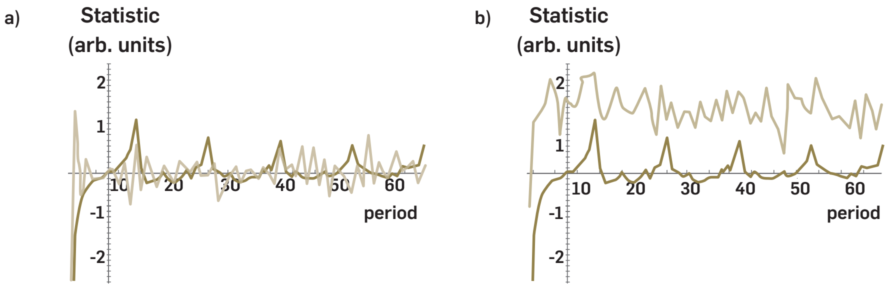
To estimate the integral in Equation (15) in the case of few parameters (5÷7, in our case), I use a suitable version of the Monte Carlo integration. Of course, this estimation method can work only if the simulated model sometimes produces paths that resemble the actually observed trajectory. Otherwise, the expression of Equations (14-15) will be random across simulation panels and numerically small[1].

Table 1 shows a preliminary list of estimation parameters. Even for a relatively small number of simulated panels (several hundred), the Student *t*-factor for the estimation is within 5% of significance for all but 1÷2 parameters. The entire Monte Carlo simulation, though, runs through 105 return-days providing a significant statistics.

[1] *I.e., $E_B[\theta] \leq Var[\theta]^{1/2}$. Practically, in that case $E_B[\theta] \sim 1/N_p^{1/2}$, where N_p is a number of panels assumed to be large.*

Given the simplicity of the model of Equations (6) and (11), this is truly remarkable. Comparison of the empirical data from HKS, Table 1, is shown in Figure 2.

Approximation of empirical data of Heston et al. (dark line) by the model of Section 5 (light line)



a) The sample with best-fitted values of parameter vector θ, r^2≈30%.
b) Sample weighted according to posterior distribution, r^2≈70%. The graph for model dynamics is shifted upwards for clarity of presentation

This estimation allows me to determine η, which is roughly the response of the market price to news (conversely, 1-η is a share of predictable information). It turns out that 30-50% of the price reaction happens in response to news and 50-70% in response to already existing information. The statistic is insufficient to distinguish this from the "half-and-half" rule, so I presume that, given the crudity of my model, approximately half of the bid-ask spread on NYSE (or any generic stock market) is formed on the basis of predictable information and the other half because of news.

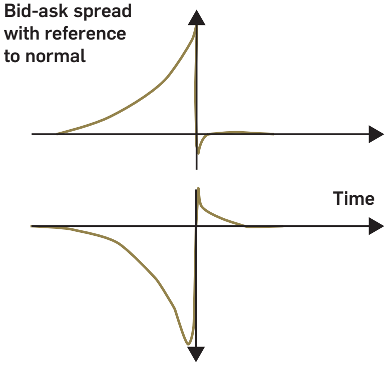
Estimated parameters of the informed trading model

σ_1	σ_2	η	$\tilde{\eta}$	$a_0=-b_0$	$\bar{V}_0=-\underline{V}_0$	υ
0.6211	0.6559	**0.3508**	**0.3402**	**0.6817**	0.6721	**0.5264**
(2.59)	(2.12)	(3.38)	(3.44)	(3.38)	(2.23)	(3.24)

To explain the bid-ask bounce of Heston et al., Equations (3) and (6) were estimated using Bayesian MLE of Equations (13-15) for the N=192 panels. Student t-coefficients are given in parentheses. The parameters significant at 1% are boldfaced.

In 2008, Madureira and Underwood produced another, quite striking, empirical demonstration of the short time scale failure of the rational expectations paradigm in event studies (Madureira and Underwood, 2008). Namely, they produced statistics of the bid-ask spread adjustment to a market event. If the rational expectations paradigm is a scientific principle rather than an article of faith, then the adjustment of the bid-ask spread to an event can only be analogous to the adjustment of the statistics of cumulative abnormal returns already established in the mid-to-late 1970s (Mandelker, 1974, Keown and Pinkerton, 1981). Namely, if an event is viewed by the market as liquidity squeeze, the bid-ask spread must build up to the time of the event and then return to its baseline value almost immediately. Conversely, if the event is viewed as liquidity enhancing, the bid-ask spread should shrink up to the time of the announcement and then jump upwards to the baseline (Figure. 3).

Hypothetical shape of the reaction of the bid-ask spread on a market event in the case of strict conformity to the efficient market hypothesis



The curves behave similarly to the derivative of the CAR function in Mandelker (1974) or Keown and Pinkerton (1981). The upper drawing shows a negative market event (liquidity squeeze); the lower drawing is a positive market event (liquidity enhancement). Practically all the action happens before the event.

Yet, what Madureira and Underwood observed was at complete variance with the rational expectations (Figure 3). First, they divided brokers into two groups: "affiliated" and "unaffiliated." The first group is affiliated with banks, which provide extensive sell-side research coverage for the stocks in question. The second group is sell-side brokers, who are not associated with research-intensive banking firms. I assume that the second group can be viewed as a proxy for the general public, which executes their orders through them.

Neither group reacted in accordance with rational expectations. First, most of the action happened *after* the moment of announcement, not *before* it as would be predicted from the efficient market hypothesis. If liquidity adjustment were a result of a pre-announcement leakage of news about the fundamental value of a company, then after the announcement there would be simply no news to leak. Furthermore, if the markets were strongly efficient there would be few differences in the behavior of research-affiliated and research non-affiliated brokers. There is no heuristic reason for information from the research analysts of investment banks to leak to the market any differently than the inside financial data of industrial firms, which comprise the index. Yet, the reaction of the two groups was strikingly different.

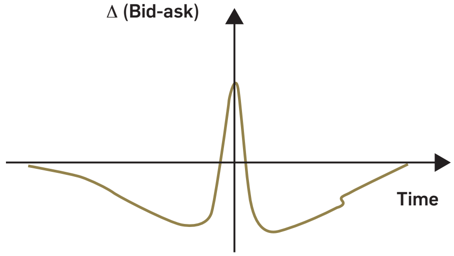
The affiliated group reacted as follows. Bid-ask spread shrank up to and through the time of the announcement, when the relative liquidity crunch was detected and then it subsided in a much deeper and longer right wing. Research-affiliated brokers increased liquidity before the event, then the liquidity squeeze spiked at the event time in a highly asymmetric fashion. Another stage of liquidity enhancement followed the market event. Research-unaffiliated brokers started to squeeze out liquidity at the time of the event and it returned to normal very slowly and in monotonic fashion.

This behavior was independent of the overall direction of liquidity in anticipation of the event. The second group adapted to the market change in a way diametrically opposite to market efficiency, in a retarded pattern: the event, the reaction and slow relaxation to a pre-event level. This behavior is analogous to the reaction to a lottery announcement: at the announcement, winning tickets are mostly removed from circulation, sharply diminishing liquidity. Hence, we can equate the reaction of the general public to market event to gambling.

The explanation of this phenomenon using microstructure concepts is quite straightforward. Liquidity, or uninformed, traders start to arrive in anticipation of the event and the crest of their arrival rate *follows* the time stamp of the announcement because they are, well, uninformed. From a practical standpoint, though, it would be a mistake to represent uninformed traders as unsophisticated. These may be large buy-side institutions such as mutual, hedge funds or insurance companies. The only assumption is that their collective behavior is statistically independent of the expected price movements and is dictated by considerations such as portfolio rebalancing, statutory requirements and servicing of contributions and redemptions.

The arrival of "uninformed" agents increases liquidity and shrinks the spread independently of whether the event itself is viewed as liquidity enhancing or reducing. Then informed traders arrive on, or around the time of the announcement and their activity, buying or selling, temporarily squeezes the liquidity out. If the actions of informed and liquidity traders were independent and the insiders were perfectly informed, the adjustment picture would be like the one shown in Figure 4. However, the majority of liquidity traders are delayed in their reaction to informed traders' behavior and the right wing of the adjustment curve is much longer and deeper than the left wing in full concordance with the observations of Madureira and Underwood.

Hypothetic reaction of the market formed by independent liquidity and informed traders to the liquidity event



Both groups of agents arrive around the event, which is accompanied by a short and almost symmetric liquidity squeeze.

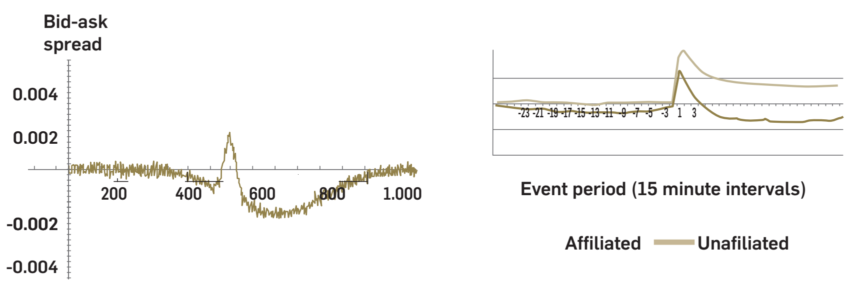
Functionally, this narrative represented in a language of EKHP equations (From Easley, Kiefer, O'Hara and Paperman, 1996), appears as follows. In the original EKHP model, the rates of arrival of liquidity and informed traders are exogenous constants, ε and μ, respectively. Near the event time, the arrival rate of liquidity as well as informed traders is modified according to the law:

$$\begin{aligned}\varepsilon(t)&=\varepsilon_0+\varepsilon_1 f_1(t)=\varepsilon_0+\frac{\varepsilon_1}{\sqrt{2\pi\sigma_u^2}}\,exp\left(-\frac{(t-t_u)^2}{2\pi\sigma_u^2}\right)\\ \mu(t)&=\mu_0+\mu_1 f_2(t)=\mu_0+\frac{\mu_1}{\sqrt{2\pi\sigma_i^2}}\,exp\left(-\frac{(t-t_i)^2}{2\pi\sigma_i^2}\right)\end{aligned} \tag{16}$$

I have chosen the Gaussian form for the distribution of trader arrivals because of its analytic simplicity but it can be proven in a number of simple probabilistic models (see Appendix B), though it is by no means unique.

The behavior of the bid-ask spread according to the above model is shown in Figure 5.

Simulations of the reaction of a bid-ask spread to a market event occurring at *T*=512 using the model of Equations (6), (11-12) with time-changing rates of arrival of traders provided by Equation (16)



We observe clear asymmetry typical for the results of Madureira and Underwood (2008). The parameters were chosen for the simulation as follows: σ_l=1E–3, σ_2=1.2E–3, $a_0=-b_0=0$, $\underline{V}_0=\overline{V}_0=0.2$, $\upsilon=1$, ε_0=0.25, ε_1=4., μ_0=0.1, μ_1=0.25, P_b=0.35, P_g=0.65. *Below, the bid-ask spread adjustment figure from the paper of Madureira and Underwood (2008) is sketched for comparison.*

While Madureira and Underwood (2008) do not provide sufficient statistics to calibrate my model from their data, the comparison of model trajectory seems to represent their empirical behavior reasonably well.

Figure 5 was simulated with the following parameters: $\sigma_1=1E{-}3$, $\sigma_2=1.2E{-}3$, $a_0=-b_0=0$, $\underline{V}_0=\overline{V}_0=0.2$, $\upsilon=1$, $\varepsilon_0=0.25$, $\varepsilon_1=4.$, $\mu_0=0.1$, $\mu_1=0.25$, $P_b=0.35$, $P_g=0.65$. All these are inessential parameters, which affect only the scale of the picture. In particular, σ_1, $\sigma_2\neq0$ are needed only to produce realistic looking "squiggles". The only parameters that influence qualitative features of the trajectory are the ones describing the time delay between the maximums of arrival of uninformed and informed traders, respectively: $(t_u-t_i)/\sigma_u=1$, and $\sigma_i/\sigma_u=0.125$. The first ratio means that the duration of delay is similar to the dispersion of arrival times of liquidity traders, and the second that the informed traders arrive and exit the market much faster than the liquidity traders.

Recently, Easley, Lopez de Prado and O'Hara (2010) produced an empirical breakthrough by defining a quantity, VPIN (volumetric probability of informed trading), which is 1) relatively easy to measure, 2) dynamically approximates the factors $\eta(t)$, $\tilde{\eta}(t)$ in the Equation (6). Moreover, the method of estimation of VPIN from the high frequency data bears a striking resemblance to the "toy" model formulated by the author in the Appendix B. Direct comparison of the two approaches is difficult for the following reason: VPIN is extremely volatile, much more so than even the high frequency returns (Easley, Prado and O'Hara, 2010, Figures 9-13). Yet, in the model of Equations (6), it is implicitly assumed that the time-dependent coefficients change slowly with respect to the prices. It is hard to see from the equations themselves but the stochastic behavior of the coefficients obviates the applicability of the Ito calculus, while Equations (3) are essentially the discrete version of Ito derivatives.

This, by itself, does not invalidate the comparison because one can always use low-pass filter (Mallat, 1999) to smooth down excess volatility of the coefficients. However, to design such a filter on the ultrashort time scale used to calculate VPIN presents a significant econometric challenge (see e.g. Lerner, 2009, Chapter 7 and *op. cit.*).

Second limitation of my approach is the simplistic modeling of the asset price process in Equation (12). A "true" asset price is a martingale only at a reasonably short time scale. In principle, the model can be appended by any asset process of choice.

In my paper, I use a modified Easley-Kiefer-O'Hara-Paperman (EKHP, 1996) model to explain two independent sets of empirical results. One set is the results of Heston, Korajczuk and Sadka (HKS, 2008) on the high-frequency bid-ask bounce of NYSE stock. Heston, Korajczuk and Sadka performed econometric analysis of a very large sample of data to rule out other possible explanations of the bid-ask bounce.

First, I develop the model into a form that can be empirically estimated. Second, I describe my method of Bayesian estimation. The data provided in Table 1 of HKS allow me to calibrate my model. The estimation is performed through the maximization of the absolute value of the averaged correlation coefficient between a single empirical trajectory of the bid-ask spread provided by HKS and a number of the MCMC-generated paths. The posterior distribution inferred by this method has $r^2 \approx 70\%$.

Another phenomenon explained through the informed trading model is the reaction to the market event. A strict constructionist interpretation of the rational expectations would predict no after-event adjustments of the bid-ask spread, but quite aggressive dynamics of the spread before the actual event. Yet, nothing of the kind was observed by Madureira and Underwood (2008). Instead, they observed that the arrival of research-affiliated brokers increases liquidity a little before the event. The announcement is followed by an asymmetric liquidity squeeze. Subsequently, we have another, longer round of liquidity enhancement.

By comparing quantitatively the bid-ask bounce and the temporary liquidity squeeze with the empirical data from the HKS and Madureira and Underwood (2008) we can hypothesize that at short- to medium time scales (from 15 minutes to 1-2 days), the price evolution is inefficient and incorporates both the new economic information and old prices.

I explain this behavior on the basis of the agent-based, or strategic trading (Hasbrouck, 2007) microstructure theory by the interaction between liquidity and informed traders. Uninformed, or liquidity, traders observing a price fluctuation have no possibility to distinguish between a real economic event and a "microstructure event", e.g. a large trade performed for hedging or regulatory reason. Only continued observation of a trading pattern eventually leads to this information being revealed (Kyle, 1985).

Liquidity traders start to arrive even before the expected event to conduct trades. Informed traders squeeze liquidity out in the immediate aftermath of a market event. The process of learning by the liquidity traders begins when they observe bids and asks by the combined trading community (liquidity + informed traders) and continues afterwards.

The group of research-unaffiliated brokers follows the pattern of asymmetric liquidity squeeze at the event time and very slow relaxation of the bid-ask spread to the pre-event levels. This picture is not symptomatic of a rational reaction but is rather reminiscent of gambling behavior when lottery tickets quickly lose liquidity after a playing round.

I thank the anonymous referees for the suggestions, which improved my paper, Prof. O'Hara for presenting and making available her working paper on measuring flow toxicity (Easley, Prado, O'Hara, 2010) for the Finance Seminar at Cornell University and Prof. J. M. Montero Lorenzo for deciding to publish such an offbeat work in his journal.

Candy, J. (2009). *Bayesian Signal Processing: Classical, Modern and Particle Filtering Method,* Wiley-Interscience, New York, NY.

EKHP. Easley, D., Kiefer, N. M., O'Hara, M. and Paperman, J. B. (1996). Liquidity, information and infrequently traded stocks, *Journal of Finance,* **51**, pp. 1405-1436.

Easley, D., Lopez de Prado, M. M. and O'Hara, M. (2010). Measuring flow toxicity in a high frequency world, Cornell University Working Paper.

Glosten, L. and Milgrom, P. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics,* **14,** pp. 71-100.

Granger, C. and Lin, J. L. (1994). Using the mutual information coefficient to identify lags in non-linear models, *Journal of Time Series Analysis*, **15**, pp. 371-384.

Harris, L. (2002). *Trading and Exchanges: Market Microstructure for Practitioners*, Oxford University Press, New York.

Hasbrouck, J. (2007). *Empirical Market Microstructure,* Oxford University Press, New York.

Heston, S., Korajczuk, R. A. and Sadka, R. (2010). Intraday Patterns in the Cross-Section of Stock Returns, *Journal of Finance,* **65**, pp. 1369-1407.

Hong, Y. (2006). *Nonlinear Time Series Analysis: Econometric Theory and Application,* Cornell University Lecture Notes.

Keown, A. and Pinkerton, J. (1981). Merger announcements and insider trading activity: an empirical investigation, *Journal of Finance,* **36**, pp. 855-869.

Kyle, A. S. (1985). Continuous auctions and insider trading, *Econometrica,* **53**, pp. 1315-1336.

Lawler, G. K. (2006). *Introduction to Stochastic Processes*, Chapman & Hall/CRC, Boca Raton, FL.

Lerner, P. (2008). Attempts at pricing of the regulatory commodity: EU carbon permits, [WP], available at http://www.ssrn.com/abstract=1426592.

Lerner, P. (2009). *Microstructure and noise in financial markets*, VDM Verlag, Saarbrucken, Germany.

Madureira, L. and Underwood, S. (2008). Information, sell-side research, and market making, *Journal of Financial Economics,* **90**, pp. 105-126.

Mallat, S. (1999). *Wavelet Tour in Signal Processing,* Academic Press, San Diego.

Mandelker, G. (1974). Risk and return: the case of merging firms, *Journal of Financial Economics,* **1**, pp. 303-335.

Plott, C. and Smith, V. (eds.) (2008). *Handbook of Experimental Economics Results,* North Holland, Amsterdam.

Derivation of the EKHP equations

The value of the asset conditional on good news on day i is $\overline{V}_i$ and conditional on bad news is $\underline{V}_i$. On any day, uninformed buyers and uninformed sellers both arrive at a rate ε. On the day with an informational event, informed traders also arrive. The arrival rate for this process is μ. Thus, on good days, the arrival rate is $\varepsilon+\mu$ for buyers and ε for sellers. On bad days, the sellers dominate buyers at the rate $\varepsilon+\mu$ to ε. The market maker determines the price given the information on the total order flow and trading history.

Let $P(t)=(P_n(t), P_b(t), P_g(t))$ be the market maker's prior belief about the events "no news" (n), "bad news" (b), and "good news" (g) at time t. Let S_t denote the event of an arrival of a "sell" order and B_t the event of a "buy" order arriving at time t.

The offer price at time 1 is then equal to

$$a_1 = E[V|B_1] = \underline{V}\,Pr\{V=\underline{V}|B_1\} + \bar{V}Pr\{V=\bar{V}|B_1\} + V_0 Pr\{V=V_0|B_1\} \tag{A.1}$$

and the bid price is

$$b_1 = E[V|S_1] = \underline{V}\,Pr\{V=\underline{V}|S_1\} + \bar{V}Pr\{V=\bar{V}|S_1\} + V_0 Pr\{V=V_0|S_1\} \tag{A.2}$$

The Bayesian update at time t=1 for the buy order yields the following formula:

$$Pr\{V=\underline{V}|B_1\} =$$

$$\frac{Pr\{V=\underline{V}\}Pr\{B_1|V=\underline{V}\}}{Pr\{V=\underline{V}\}Pr\{B_1|V=\underline{V}\}+Pr\{V=\bar{V}\}Pr\{B_1|V=\bar{V}\}+Pr\{V=V_0\}Pr\{B_1|V=V_0\}} \tag{A.3}$$

The posterior probability is equal to

$$P_b(1|B_1) = \frac{P_b(0)\varepsilon}{\varepsilon + P_g(0)\mu} \tag{A.4}$$

Similarly, all other probabilities are computed. Substituting them into the formula for the ask price, we have:

$$a_1 = \frac{P_n(0)\varepsilon V_0 + P_b(0)\varepsilon \underline{V} + P_g(0)(\varepsilon+\mu)\bar{V}}{\varepsilon + P_g(0)\mu} \tag{A.5}$$

For the bid price, the following formula is valid:

$$b_1 = \frac{P_n(0)\varepsilon V_0 + P_b(0)(\varepsilon+\mu)\underline{V} + P_g(0)\varepsilon \bar{V}}{\varepsilon + P_b(0)\mu} \tag{A.6}$$

One can express an expected value prior to the trade at time t=1:

$$E[V|t<1] = P_n(0)V_0 + P_b(0)\underline{V} + P_g(0)\bar{V} \tag{A.7}$$

Hence, we obtain

$$a_1 = E(V_1|0) + \frac{\mu P_{g,0}}{\varepsilon + \mu P_{g,0}}(\bar{V}_1 - E(V_1|0)) \tag{A.8}$$

$$b_1 = E(V_1|0) + \frac{\mu P_{b,0}}{\varepsilon + \mu P_{b,0}}(\underline{V}_1 - E(V_1|0))$$

By induction, one can propagate these equations up to time t, thus deriving the system of Equations (2) of the main text.

The EKHP equations use only the Bayes formula for their derivation. Consequently they are very general. Their dynamic can incorporate an arbitrary trading strategy, i.e., the measure according to which one computes the expectations in (A.8), as well as an arbitrary asset process (the rules of change for $\overline{V}_t, \underline{V}_t$ and P_g, P_b with time) and the stochastic rates of arrival of informed μ and liquidity traders ε to the market. The last opportunity was already explored in the first paper on the subject, EKHP (1996). However, with a few simplifying assumptions, one can cook up a continuous-time market learning theory, which can be exactly solved and has attractive features, such as convergence of the expectations of liquidity traders to the true price.

If we substitute ask and bid prices at time t into Equation (A.8), we obtain:

$$a_{t+1}=V_t\left(1-\frac{\mu P_g}{\varepsilon+\mu P_g}\right)^2+\left(\frac{2\mu P_g}{\varepsilon+\mu P_g}-\left(\frac{\mu P_g}{\varepsilon+\mu P_g}\right)^2\right)\overline{V}_t+\frac{\mu P_g}{\varepsilon+\mu P_g}u_{t+1}$$

$$b_{t+1}=V_t\left(1-\frac{\mu P_b}{\varepsilon+\mu P_b}\right)^2+\left(\frac{2\mu P_b}{\varepsilon+\mu P_b}-\left(\frac{\mu P_b}{\varepsilon+\mu P_b}\right)^2\right)\underline{V}_t+\frac{\mu P_b}{\varepsilon+\mu P_b}v_{t+1} \quad \text{(A.9)}$$

where

$$u_{t+1}=\overline{V}_{t+1}-\overline{V}_t, \quad v_{t+1}=\underline{V}_{t+1}-\underline{V}_t, \quad \text{(A.10)}$$

are the asset price innovations at time $t+1$. Intuitively, we can identify the first coefficient in each of the Equations (A.9) with the probability of the "stale" prices, the second coefficient with the probability of the asset price movement and the third coefficient with the strength of the microstructure noise. From now on we consider the frequency of the good and bad events as exogenously given constants. From Equations (A.9), the expectations of the bid and ask prices conditional on the price history observed up to time t are

$$E(a_{t+1}|t)=V_t(1-\eta)^2+(2\eta-\eta^2)\,\overline{V}_t$$

$$E(b_{t+1}|t)=V_t(1-\tilde{\eta})^2+(2\tilde{\eta}-\tilde{\eta}^2)\,\underline{V}_t \quad \text{(A.11)}$$

because and $E(u_{t+1}\,|\,t)=0$, $E(v_{t+1}\,|\,t)=0$. [2] In (A.11),

$$\eta=\frac{\mu P_g}{\varepsilon+\mu P_g}, \qquad \tilde{\eta}=\frac{\mu P_b}{\varepsilon+\mu P_b}$$

[2] *In statistical parlance microstructure noise is represented by MDS (Martingale Difference Sequences).*

Descriptive example of a model with a Gaussian envelope

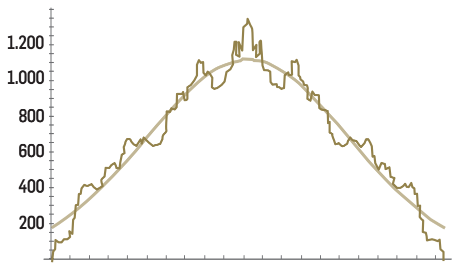
Because of the Central Limit Theorem the emergence of Gaussian distributions in statistics is ubiquitous. However, examples when the envelope of a statistical function has a Gaussian profile in time, though numerous, are much harder to find outside of physics. In this Appendix, I describe one possible model in the trading context.

We have $N/2$ portfolios (N is large) of securities/commodities: "the short buckets," which are filled by the buy/sell orders coming according to a Poisson distribution with a random frequency obeying some statistical distribution. This may be, for instance, a uniform distribution between 0 and 1. When the bucket is filled, there is no more activity in a given stock.

A similar set of buckets is set for the long portfolios, namely they are depleted by the buy/sell orders also coming with a random frequency chosen separately for each of the $N/2$ portfolios in a Poisson distribution. Activity ends when there are zero stocks in the portfolio. Then, in the compound "market," the average number of orders per unit time has an approximately Gaussian shape.

The result of the model described above for N=480 portfolios and T=2048 moments of time is shown in Figure 6 together with the Gaussian approximation of the envelope.

Plot of the envelope of the trading volume for the model being described in the Appendix B



The plot of the envelope of the trading volume for the model being described in the Appendix B. Best fit by the Gaussian curve is shown as a guide to the eye. We plot the number of orders on the vertical axis as a function of the moments of time on the horizontal axis.