PHYSICAL REVIEW E **71**, 046131 (2005)

Quantifying fluctuations in market liquidity: Analysis of the bid-ask spread

Vasiliki Plerou,* Parameswaran Gopikrishnan,† and H. Eugene Stanley‡
Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA
(Received 15 March 2004; published 22 April 2005)

Quantifying the statistical features of the bid-ask spread offers the possibility of understanding some aspects of market liquidity. Using quote data for the 116 most frequently traded stocks on the New York Stock Exchange over the two-year period 1994–1995, we analyze the fluctuations of the average bid-ask spread S over a time interval Δt. We find that S is characterized by a distribution that decays as a power law $P\{S>x\}\sim x^{-\zeta_S}$, with an exponent $\zeta_S\approx 3$ for all 116 stocks analyzed. Our analysis of the autocorrelation function of S shows long-range power-law correlations, $\langle S(t)S(t+\tau)\rangle\sim\tau^{-\mu_s}$, similar to those previously found for the volatility. We next examine the relationship between the bid-ask spread and the volume Q, and find that $S\sim\ln Q$; we find that a similar logarithmic relationship holds between the transaction-level bid-ask spread and the trade size. We then study the relationship between S and other indicators of market liquidity such as the frequency of trades N and the frequency of quote updates U, and find $S\sim\ln N$ and $S\sim\ln U$. Lastly, we show that the bid-ask spread and the volatility are also related logarithmically.

DOI: 10.1103/PhysRevE.71.046131 PACS number(s): 89.90.+n, 05.45.Tp, 05.40.−a, 05.40.Fb

I. INTRODUCTION

The primary function of a market is to provide a venue where buyers and sellers can transact. The more buyers and sellers at any time, the more efficient the market is in matching buyers and sellers, so a desirable feature of a competitive market is liquidity, i.e., the ability to transact quickly with small price impact. To this end, most exchanges have market makers [e.g., "specialists" in the New York Stock Exchange (NYSE)] who provide liquidity by selling or buying according to the prevalent market demand. The market maker sells at the "ask" (offer) price A and buys at a lower "bid" price B; the difference $s\equiv A-B$ is the bid-ask spread.

The ability to buy at a low price and sell at a high price is the main compensation to market makers for the risk they incur while providing liquidity. Therefore, the spread must cover costs incurred by the market maker [1–7] such as: (i) order processing costs, e.g., costs incurred in setting up, fixed exchange fees, etc., (ii) risk of holding inventory, which is related to the volatility, and (iii) adverse information costs, i.e, the risk of trading with a counterparty with superior information. Since the first component is a fixed cost, the interesting dynamics of liquidity is reflected in (ii) and (iii). Analyzing the statistical features of the bid-ask spread thus also provides a way to understand information flow in the market.

The prevalent bid-ask spread reflects the underlying liquidity for a particular stock. Quantifying the fluctuations of the bid-ask spread thus offers a way of understanding the dynamics of market liquidity. In this paper, we show that the fluctuations of the average bid-ask spread over a fixed time interval display power-law distributions and long-range correlations. We further explore the relationship between the bid-ask spread and other indicators of liquidity such as the frequency of trade occurrence N, and the frequency of quote updates U. We find $S\sim\ln N$ and $S\sim\ln U$. We find a similar logarithmic relation between the bid-ask spread and the share volume, both over a fixed time interval and on a transaction level. Lastly, we report logarithmic relationships between bid-ask spread, order flow, and two different measures of volatility.

Our analysis focuses on stocks that are listed on the NYSE. The NYSE is a hybrid market in which both the specialist and limit-order traders play a role in price formation. The hybrid market system ensures that specialists incorporate the best bid and ask in the order book while posting their quotes. The NYSE hybrid market contrasts with a purely order-driven market, such as the Tokyo Stock Exchange, where orders are submitted before prices are determined, or a "quote-driven" system such as used in NASDAQ, where different competing market makers are required to provide bid-ask quotes continuously.

In an order-driven market, orders are submitted to a centralized location (electronic or physical), where they are matched, executed, or deleted. Here, the bid price represents the largest sell limit order price and the ask price represents the smallest buy limit order price. Their difference defines the spread. Order-driven markets are generally cheaper to trade since they have smaller bid-ask spreads, in part because fixed costs such as (i) discussed above, are not present.

We analyze the trades and quotes (TAQ) database for the two-year period January 1994 to December 1995. The TAQ database, which has been published by the NYSE since 1993, covers all trades and quotes for all stocks listed at three major U.S. stock markets (NYSE, AMEX, and NASDAQ). Our analysis focuses on a subset of these stocks that are traded on the NYSE.

This paper is organized as follows. Sections II and III present our results on the distribution and time correlations of the bid-ask spread, respectively. Section IV presents our results on the relationship between the bid-ask spread, the

*Corresponding author. Email address: plerou@bu.edu
†Present address: Goldman Sachs & Co., 85 Broad Street, New York, NY 10004.
‡Email address: hes@bu.edu

share volume, and the number of trades, both over fixed time intervals and at the transaction level. Sections V and VI describe the relationship of the bid-ask spread to the order flow and to the frequency of quote updates. Section VII explores the relationship between the bid-ask spread and the volatility.

II. DISTRIBUTION OF FLUCTUATIONS IN THE BID-ASK SPREAD

From the TAQ database, we select the 116 most frequently traded stocks listed in the NYSE, and form for each stock a list of all trades and quotes. For each trade i, the database provides the trade price P_i and the trade size q_i. We use the procedure of Ref. [8] to identify for each trade i, the bid price B_i and the ask price A_i.

We first compute a time series of the average spread S

$$S \equiv S_{\Delta t}(t) \equiv \frac{1}{N}\sum_{i=1}^{N} s_i, \tag{1}$$

over fixed time intervals Δt, where $s_i \equiv A_i - B_i$ [9] and $N \equiv N_{\Delta t}(t)$ denotes the number of trades in Δt [10]. In the following analysis, we show results for $\Delta t = 15$ min. We find similar results for $\Delta t = 30$ and 60 min.

Denote by $S_j(t)$ the time series of the bid-ask spread for stock j. Figure 1(a) shows that the $S_j(t)$ for a typical stock displays large fluctuations. We compute the cumulative distribution function $P\{S_j > x\}$ for each of the $j = 1, \ldots, 116$ stocks. We find that each distribution is consistent with a power law

$$P\{S_j > x\} \sim x^{-\zeta_{S_j}}. \tag{2}$$

Our estimates of the individual exponents ζ_{S_j} are similar across all 116 stocks in our sample. Although the functional forms of individual distributions are similar, their widths (standard deviations) vary. We obtain a good "collapse" of the distributions by normalizing each time series $S_j(t)$ by transforming it to zero mean and unit standard deviation. Based on the hypothesis that the functional forms of the distributions $P\{S_j > x\}$ are the same for all stocks, we compute the cumulative distribution function $P\{S > x\}$ using the normalized spreads for all stocks $j = 1, \ldots, 116$. Figure 1(b) shows that $P\{S > x\}$ decays as

$$P\{S > x\} \sim x^{-\zeta_S}, \tag{3}$$

and we find the mean value $\zeta_S = 3.0 \pm 0.1$.

We note that ζ_S is similar in value to the exponent found for ζ_G describing the distribution $P\{G > x\}$ of price change G [11,12], and the tail exponent describing the distribution of volatility [13]. Our analysis of the relationship between return G and S shows an approximate power-law dependence $|G| \sim S^{\alpha}$ with $\alpha \approx 0.7$. For most stocks this dependence holds only up to a threshold, after which there is a drop off. Since the relationship is weaker than linear ($\alpha < 1$ followed by a drop off), our results do not seem to support the simplistic hypothesis that the power-law tails of $|G|$ with $\zeta_G \approx 3$ can be explained by $\zeta_S \approx 3$.

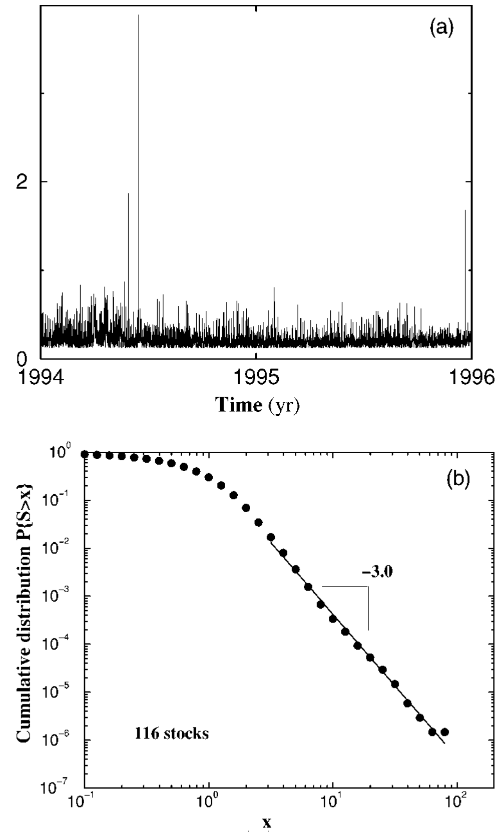

FIG. 1. (a) Time series of the bid-ask spread over $\Delta t = 15$ min for a typical stock, Exxon Corp., for the two-year period 1994–1995. The smallest value of the abscissa is the tick size for this stock $= 1/8$ USD. (b) The log-log plot of the cumulative distribution of $S_{\Delta t}$, which is normalized to have a zero mean and unit variance, for all 116 stocks in our sample. The abscissa is therefore in units of standard deviations. A power-law fit in the region $x > 3$ gives a value for the exponent $\zeta_S = 3.0 \pm 0.1$. Fits to individual distributions give similar results for the exponent values.

III. TIME CORRELATIONS IN THE BID-ASK SPREAD

We next consider temporal correlations in the bid-ask spread. Figure 2(a) shows the autocorrelation function $\langle S(t)S(t+\tau)\rangle$ for a typical stock, where $S(t)$ is transformed to zero mean and unit variance. We find that $\langle S(t)S(t+\tau)\rangle$ decays slowly and displays pronounced peaks at multiples of 1 day (390 min). The peaks originate from the U-shaped intraday pattern in the bid-ask spread [14], similar to the previously reported intradaily patterns in volatility [13,15–17].

To test the presence of long-range correlations, we first remove the intraday pattern from $S(t)$ using the procedure outlined in Ref. [13]. To accurately quantify the long-range persistence of the bid-ask spread correlation function, we use the method of detrended fluctuation analysis (DFA). We therefore calculate the detrended fluctuation function $F(\tau)$, defined as the *root-mean-square* fluctuation around a polynomial fit to the integrated time series of S (for details see Ref. [13]). In the analysis presented in this paper we use linear detrending. We find that the detrended fluctuation function $F(\tau)$ for S scales as

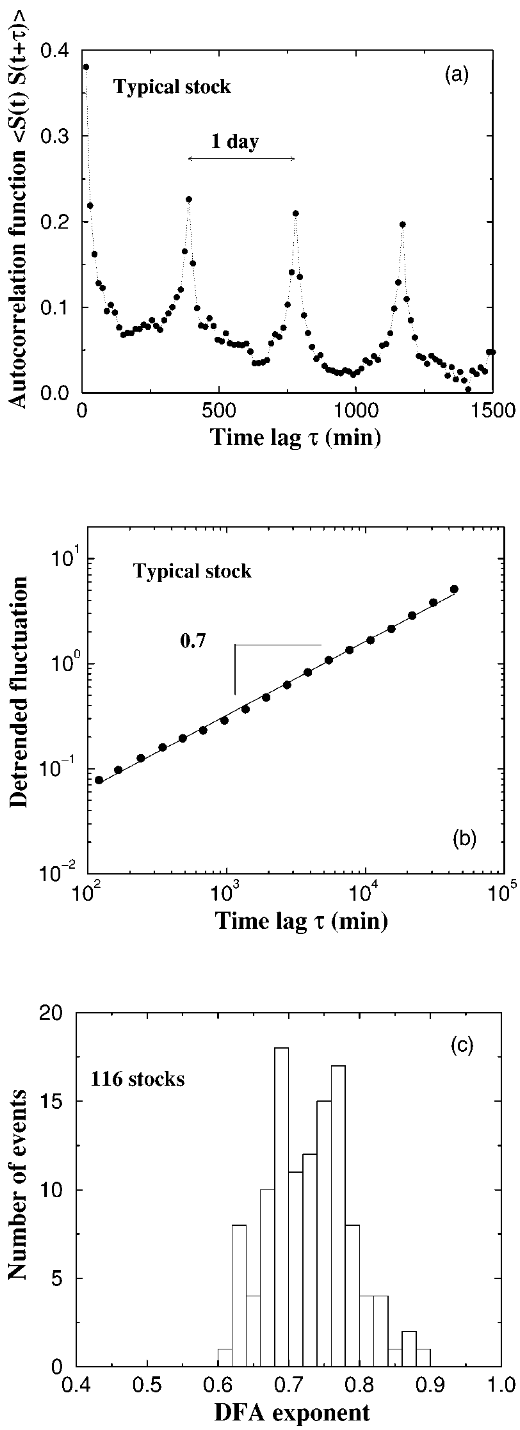

FIG. 2. (a) The autocorrelation function $\langle S(t)S(t+\tau)\rangle$ displays peaks at multiples of one day for Exxon Corp. (b) The detrended fluctuation function $F(\tau)$ for the same stock displays long-range power-law correlations that extend over almost 3 orders of magnitude. (c) A histogram of slopes are obtained by fitting $F(\tau)=\tau^{\nu_S}$ for all 116 stocks. We find a mean value of the exponent $\nu_S=0.73\pm0.01$. The error bar denotes the standard error of the mean of the distribution of exponents, which, under *i.i.d.* assumptions, is estimated as the ratio of the standard deviation of the distribution to the square root of the number of points. In reality, the *i.i.d.* assumptions do not hold, so the error bar thus obtained is likely understated.

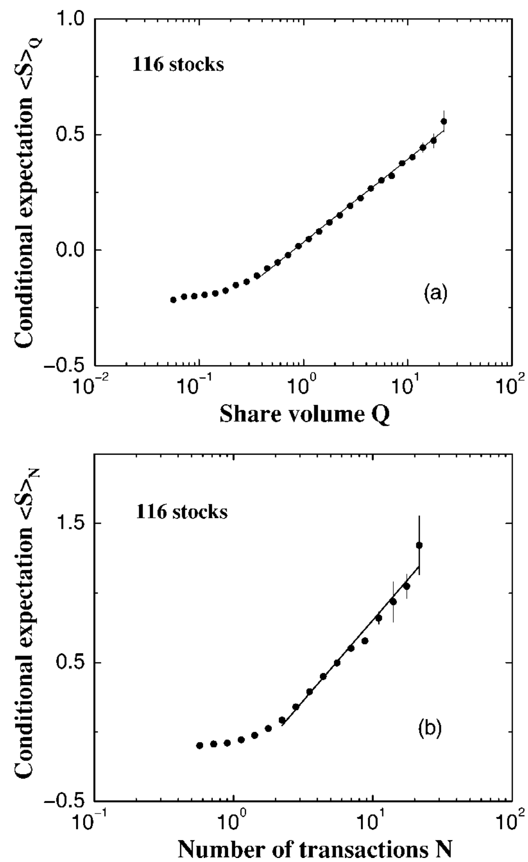

FIG. 3. (a) Equal-time conditional expectation $\langle S\rangle_Q$ of the spread for a given value of Q averaged over all 116 stocks over a time interval $\Delta t=15$ min. Here S is normalized to have a zero mean and unit variance, and $Q\equiv Q_{\Delta t}(t)$ is normalized by its first centered moment. The solid line shows a logarithmic fit to the data extending over almost two orders of magnitude. (b) Conditional expectation $\langle S\rangle_N$. As before, S is normalized to have a zero mean and unit variance, and we normalize $N\equiv N_{\Delta t}(t)$ by its standard deviation. The solid line shows a logarithmic fit to the data. Note that for both (a) and (b) the ordinate takes negative values because of our normalization of the spread to zero mean. We have tested that these relationships are robust under other normalization schemes such as scaling by the first moment.

$$F(\tau) \sim \tau^{\nu_S}, \tag{4}$$

with the mean value of $\nu_S=0.73\pm0.01$ for all 116 stocks [Figs. 2(b) and 2(c)]. The correlation function therefore decays as

$$\langle S(t)S(t+\tau)\rangle \sim \tau^{-\mu_S}, \tag{5}$$

with $\mu_S=2-2\nu_S=0.54\pm0.02$.

The power-law distributions and long-range correlations that we find in S are similar to those found in the volatility (measured, e.g., by $|G|$) [13,15–18]. The similarity in statistical properties of spread and volatility is qualitatively consistent with the notion that spreads reflect the market maker's risk of holding inventory, which is, in turn, an increasing function of volatility [19].

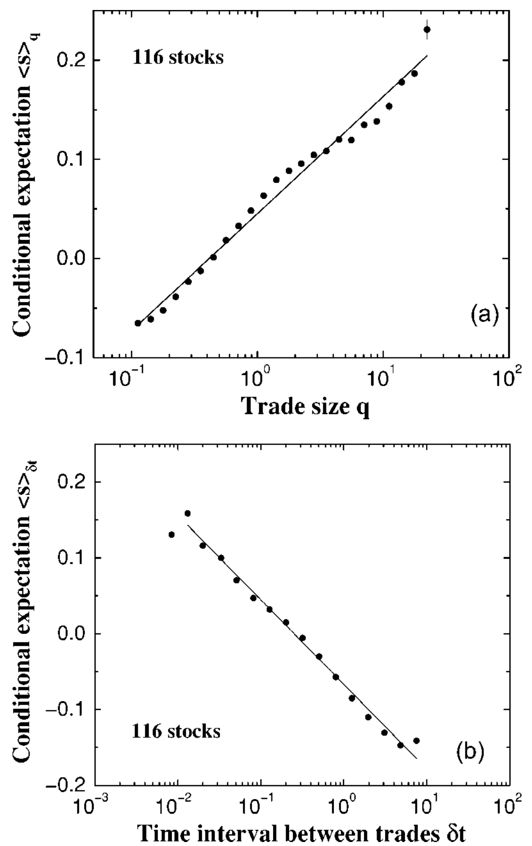

FIG. 4. (a) Conditional expectation of the transaction-level spread conditioned on trade size averaged over all 116 stocks. The solid line is a logarithmic fit to the data. Here s has been normalized to have a zero mean and unit variance, and q is normalized by its centered first moment. For trade i, instead of conditioning solely on its size q_i, we have taken a local average of q_i over four preceding trades to account for the rapidly decaying correlation function $\langle q_i s_{i+k} \rangle$. The logarithmic result also holds without the local averaging. (b) The expectation of the spread conditioned on the time interval δt between trades. The solid line shows a logarithmic fit to the data. Here δt is normalized by its standard deviation.

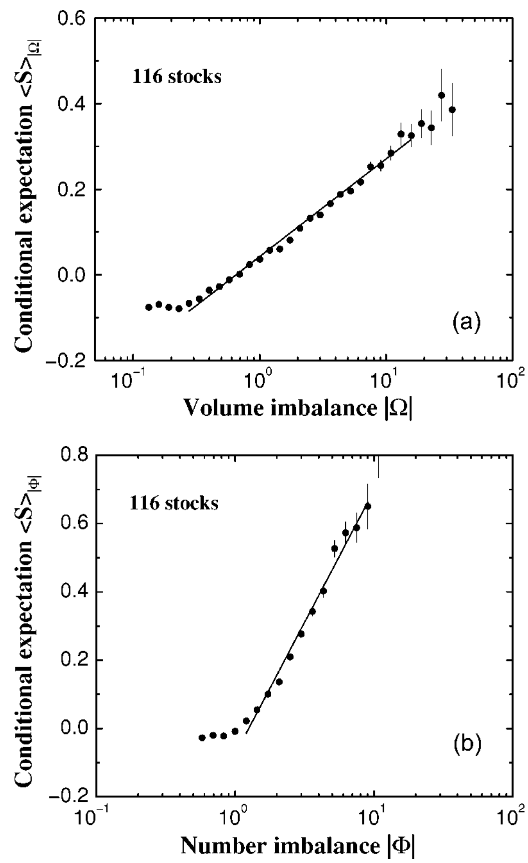

FIG. 5. (a) Conditional expectation $\langle S \rangle_{|\Omega|}$ of the spread for a given value of $|\Omega|$ averaged for all 116 stocks over a time interval $\Delta t = 15$ min. The solid line shows a logarithmic fit to the data. Here, Ω is normalized by its first moment after setting to zero mean. (b) The conditional expectation $\langle S \rangle_{|\Phi|}$. Here, Φ is normalized to a zero mean and unit variance. In both plots, S is normalized by setting to a zero mean and unit variance.

IV. BID-ASK SPREAD, SHARE VOLUME, AND TRADING ACTIVITY

A. Fixed time interval analysis

We next examine the relationship between the bid-ask spread $\langle S \rangle_Q$ and the share volume Q traded [20]. Both S and Q display intraday patterns and are large near the open and close of the market and smaller around midday. Figure 3(a) shows that the increase of spread with volume is consistent, over 2 orders of magnitude, with the logarithmic relationship [21]

$$\langle S \rangle_Q \sim \ln Q. \tag{6}$$

One may expect that small spreads should accompany large volumes, since one expects that counterparties are easier to find during times of large activity. Here we find the opposite relation, i.e., a positive correlation between the equal time conditional expectation $\langle S \rangle_Q$ and Q. Perhaps the reason for this increasing relationship is that large volumes tend to be directional (buy or sell), so they consume prevalent orders in the order book, thereby increasing the spread. If so, the logarithmic relation that we find, particularly at the transaction level (below), reflects the distribution of the order book [22,23].

We next analyze the relationship between $S_{\Delta t}(t)$ and the number of trades $N_{\Delta t}(t)$. Figure 3(b) shows that the increase with N of the equal-time conditional expectation $\langle S \rangle_N$ can be fit by a logarithmic function

$$\langle S \rangle_N \sim \ln N. \tag{7}$$

For both Q and N, we test and confirm that the logarithmic relationships hold individually for each stock.

Recent empirical studies [17] show that the long-range correlations in volatility V and volume Q can be related to the long-range correlations in N. This is because $V \sim \sqrt{N}$ and $Q \sim N$, and N has recently been shown to be long-range correlated [17,20]. Similarly, since $\langle S \rangle_N \sim \ln N$, it is not surprising that the long-range correlation in S also arises from the long-range correlations in N.

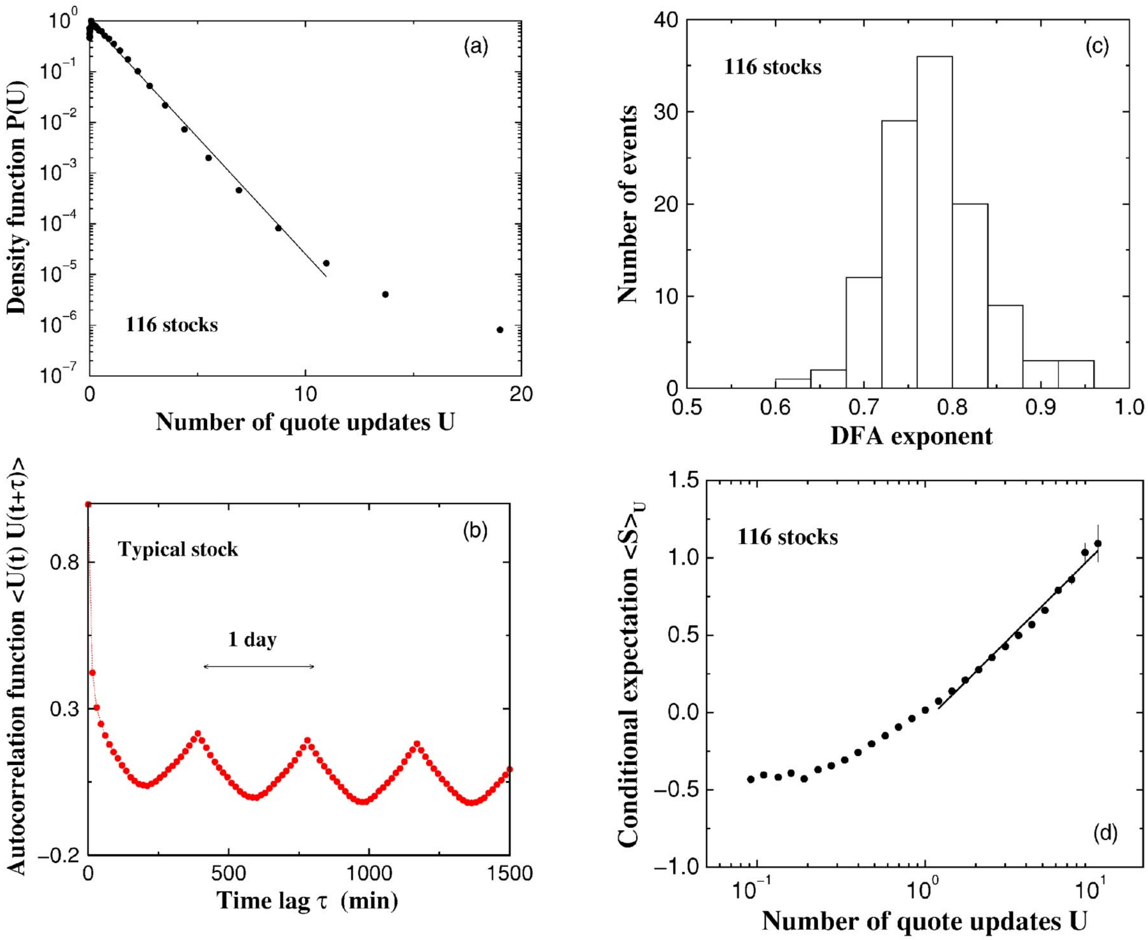

FIG. 6. (a) Probability density function $P(U)$ for the number of quote revisions U for all 116 stocks. Here U for each stock is normalized to a zero mean and unit variance. (b) The correlation function $\langle U(t)U(t+\tau)\rangle$ shows long-range correlations (with a marked intraday pattern). (c) The histogram of our estimates of DFA exponents for each stock. We find a mean exponent value of $\nu_u=0.78\pm0.03$. As before, the error bar denotes the standard error of the mean. (d) The conditional expectation $\langle S\rangle_U$ of the spread for a given quote-update frequency. The solid line shows a logarithmic fit to the data. Here U has been normalized by its second moment.

B. Transaction level analysis

We next analyze the relationship between spread and volume at the trade level. To test the time dependence between spreads and volume, we first analyze the correlation function $\langle q_i s_{i+k}\rangle$. The correlation function has its largest value at $k=0$; for $k<0$, correlations are almost zero while for $k>0$ we find correlations that decay to zero quickly. Beyond $k=4$ trades, we find no statistically significant correlation.

We next analyze the conditional expectation $\langle s\rangle_q$ of the transaction-level bid-ask spread conditioned on the trade size. We find [Fig. 4(a)], similar to Eq. (6),

$$\langle s\rangle_q \sim \ln q. \tag{8}$$

The logarithmic relationship $S\sim\ln Q$ [Eq. (6)] can therefore be understood because $S=1/N\Sigma_1^N s_i$, and $\Sigma_{i=1}^N s_i=\Sigma_{i=1}^N \ln q_i$ can be expressed to leading order in terms of $\ln(\Sigma_{i=1}^N q_i)=\ln Q$, and consequently $S\sim\ln Q$.

We next examine the relationship between s_i and the time interval δt between trades. The average intertrade time interval $\langle\delta t\rangle$ can be thought of as a reciprocal of $N_{\Delta t}$ to a first approximation. We find that as δt *increases*, the bid-ask spread *decreases*, and the functional relationship is [Fig. 4(b)],

$$\langle s\rangle_{\delta t} \sim -\ln\,\delta t, \tag{9}$$

where the brackets denote an average over all transactions i conditioned on δt.

V. SPREADS AND ORDER FLOW

Similar logarithmic functional forms also describe the relation between the bid-offer spread and order flow. During periods of large demand or supply, we expect S to be large, since a market maker increases the spread to compensate for the additional risk.

Denote $a_i=-1$ if the trade is seller initiated and $a_i=1$ if the trade is buyer initiated. The volume imbalance can then be defined as [24]

$$\Omega_{\Delta t}(t) \equiv \Omega \equiv \sum_{i=1}^{N} q_i a_i, \tag{10}$$

and the number imbalance

$$\Phi \equiv \sum_{i=1}^{N} a_i, \tag{11}$$

which quantify, respectively, the difference between the trading volume and number [25] of buyer-initiated and seller-initiated trades in a time interval Δt. We next compute the equal-time conditional expectations $\langle S\rangle_{|\Omega|}$ [Fig. 5(a)] and $\langle S\rangle_{|\Phi|}$ [Fig. 5(b)] of the spread for a given value of volume imbalance and number imbalance, respectively. We find

$$\langle S\rangle_{|\Omega|} \sim \ln(|\Omega|). \tag{12}$$

and

$$\langle S\rangle_{|\Phi|} \sim \ln(|\Phi|). \tag{13}$$

The logarithmic relationship $\langle S\rangle_{|\Omega|} \sim \ln(|\Omega|)$ is not surprising, since $|\Omega| \sim Q$ for large volumes, and we have seen above that $S \sim \ln Q$.

VI. FREQUENCY OF QUOTE UPDATES

Our analysis thus far has focused on the properties of the bid-offer spread. A closely related indicator of liquidity is the quote-update frequency $U \equiv U_{\Delta t}(t)$, i.e., the number of times a new bid or offer is posted in the market in a time interval Δt. Note that the prevalent bid or offer can change either because of incoming market orders, limit-order cancellations, or by the specialist posting an improved quote over the prevalent best limit-order book bid or ask. We therefore analyze the statistics of U to understand the behavior of liquidity in terms of quote updates.

Figure 6(a) shows that the distribution $P(U)$ decays almost exponentially, unlike the frequency of trades N, which has power-law fluctuations [17]. Performing power-law fits gives a very large value of exponent, consistent with an approximately exponential behavior.

We next consider temporal correlations in the quote-update frequency. Figure 6(b) shows the autocorrelation function $\langle U(t)U(t+\tau)\rangle$ for a typical stock where $U(t)$ is transformed to zero mean and unit variance. As before, we find that $\langle U(t)U(t+\tau)\rangle$ decays slowly and displays pronounced peaks at multiples of 1 day (390 min), similar to the intraday pattern that we find in S. To accurately quantify these correlations, we use the DFA method and find that the de-trended fluctuation function for U scales as

$$F_U(\tau) \sim \tau^{\nu_u}, \tag{14}$$

with the mean value $\nu_u = 0.78 \pm 0.03$ for all 116 stocks [Fig. 6(c)]. Here, we have first excluded the effects of the intraday pattern, and have performed linear detrending for computing $F_U(\tau)$.

The correlation function $\langle U(t)U(t+\tau)\rangle$ correspondingly decays as

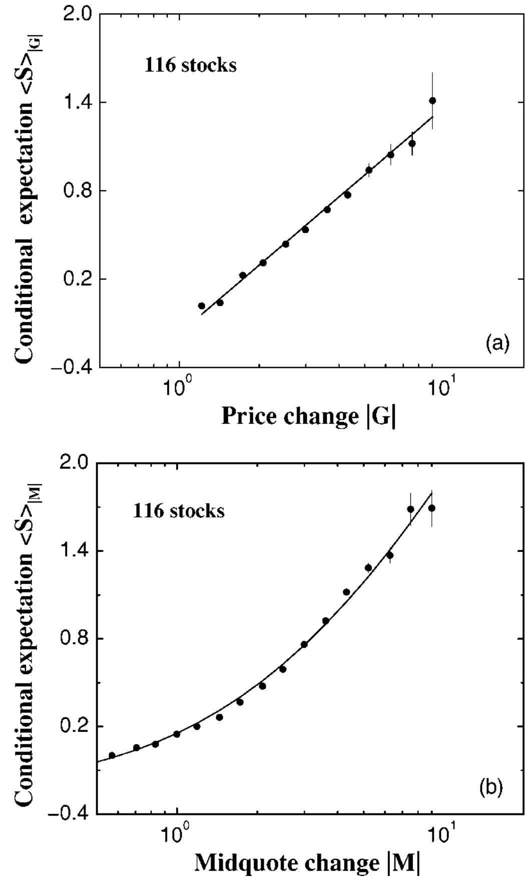

FIG. 7. (a) Conditional expectation $\langle S\rangle_{|G|}$ of the spread for a given volatility (estimated by $|G|$) averaged over all 116 stocks. The solid line shows a logarithmic fit. (b) The conditional expectation $\langle S\rangle_{|M|}$ of the spread for a given magnitude of the midquote change $|M|$. The fit is curved because we use a shifted logarithmic fit $A + B\ \ln(x+x_0)$ to the data. A reasonable fit can also be obtained by a power law; we find small exponent values. In (a) and (b) G, S, and M have been normalized to have a zero mean and unit variance.

$$\langle U(t)U(t+\tau)\rangle \sim \tau^{-\mu_u}, \tag{15}$$

with $\mu_u = 2 - 2\nu_u = 0.44 \pm 0.06$. Somewhat related results are obtained in Ref. [26].

As we have found previously, spreads depend logarithmically on the number of trades. Similarly, we find a logarithmic relationship between spreads and the frequency of quote revisions. Figure 7(d) shows that the conditional expectation

$$\langle S\rangle_U \sim \ln U. \tag{16}$$

VII. RELATION BETWEEN SPREADS AND VOLATILITY

Finally, we study the dependence of the bid-ask spread on the volatility of price movements. As a short-term estimate of volatility, we consider two measures: (a) the magnitude $|G| \equiv |G|_{\Delta t}(t)$ of the price changes and (b) the magnitude $|M| \equiv |M|_{\Delta t}(t)$ of the midquote change over the time interval

Δt. Figure 6(a) shows that the equal-time conditional expectation $\langle S\rangle_{|G|}$ of the spread for a given value of $|G|$ increases logarithmically,

$$\langle S\rangle_{|G|} \sim \ln(|G|). \tag{17}$$

We find similar behavior for the spread when conditioned on the absolute changes in the midquote price $|M|$ [Fig. 6(b)],

$$\langle S\rangle_{|M|} \sim \ln(|M|). \tag{18}$$

It is interesting to contrast our finding of a logarithmic dependence of the bid-ask spread on the volatility with the somewhat related results of Ref. [27], where an almost linear relationship is reported between the time-averaged spread and the time-averaged volatility for LSE stocks. While the linear relationship of Ref. [27] is a time-averaged property that holds between the mean spread and the mean volatility for a particular stock, our finding of Eq. (18) reflects more on the dynamics of the joint evolution of $S(t)$ and $|M(t)|$.

VIII. DISCUSSION AND SUMMARY

The relationships that we uncover for the bid-ask spread are interesting from the perspective of recent work [22,23,26,28–30]. Reference [26] analyzes the Island electronic communication network (ECN) order book, which is one of many electronic platforms that comprise NASDAQ. They report the long-memory behavior of the rates of order placement and cancellation, that is related to our finding of long memory in S. Studying double-auction limit-order markets within a model where order arrivals and cancellations follow a Poisson process, Ref. [29] finds an exponential tail for the distribution of spread, and a weak approximately linear relationship between spreads and order size. It is possible that their inherent assumptions about thin-tailed distribution of order sizes or the *i.i.d.* nature of order flow gives rise to the disparity with our empirical finding of $S \sim \ln Q$.

Reference [30] studies the evolution of spreads and volatility following large price moves in the NYSE and NASDAQ. They find that for the NYSE, both the volatility and the bid-ask spread decay as a power law following a large price move. For NASDAQ, however, they find that the bid-ask spread reacts in a much milder way than for the NYSE (a 20% increase compared to a 600% increase for the NYSE [30]). On the other hand, Ref. [23] finds that their analysis of the Island ECN order book for NASDAQ stocks seems to give qualitatively similar results to their analysis of the Paris Bourse. In light of these findings, it is interesting to see if the relationships that we uncover for the spread and its relation to volume in the NYSE hold for quote-driven markets as well.

In sum, we have analyzed the statistical properties of the bid-ask spread for the most frequently traded 116 stocks in the NYSE. We have found that the bid-ask spread S over a fixed time interval Δt displays power-law distributions and long-range temporal correlations. Our finding that $\langle S\rangle_N \sim \ln N$ suggests that the long-range correlations in S arises from those of N. We have explored the relationship between the bid-ask spread and the transaction volume and find a logarithmic relationship both over fixed time intervals and at the trade-by-trade level. Lastly, we have found logarithmic relationships between spreads, order flow, and volatility. Our results add to the existing literature on the relationships between spreads and volatility, and uncover interesting logarithmic relationships that may offer a guide to modeling the microstructural dynamics of spreads, returns, volume, and volatility.

ACKNOWLEDGMENTS

We thank Xavier Gabaix for helpful discussions. We thank the NSF and the Morgan Stanley Microstructure Research Grant for support. PG's contribution to this work was primarily during his thesis work at Boston University.

[1] T. Copeland and D. Galai, J. Financ. **38**, 5 (1983).
[2] D. Easley and M. O'Hara, J. Financ. Econ. **19**, 69 (1987).
[3] R. Roll, J. Financ. **39**, 1127 (1984).
[4] J. Y. Campbell, A. Lo, and A. C. MacKinlay, *The Econometrics of Financial Markets* (Princeton University Press, Princeton, 1999).
[5] Y. Amihud and H. Mendelson, J. Financ. Econ. **8**, 31 (1980).
[6] L. Glosten and P. Milgrom, J. Financ. Econ. **14**, 71 (1985).
[7] J. Hasbrouck, J. Financ. Econ. **22**, 229 (1988).
[8] Following the procedure of C. M. Lee and M. J. Ready, J. Financ. **46**, 733 (1991), we use the prevailing quote at least 5 s prior to the trade. Lee and Ready report that 59.3% of the quotes are recorded prior to trade. They find that using the prevailing quote at least 5 s prior to the trade mitigates this problem. See also Ref. [7].
[9] We consider only those quotes that correspond to an actual trade.
[10] Two alternative definitions are (i) the difference between the highest bid and the lowest ask in Δt and (ii) the ratio of $S(t)$ to the prevalent midquote (proportional spread).
[11] T. Lux, Appl. Financ. Econ. **6**, 463 (1996).
[12] V. Plerou *et al.*, Phys. Rev. E **60**, 6519 (1999); P. Gopikrishnan *et al.*, *ibid.* **60**, 5305 (1999).
[13] Y. Liu *et al.*, Phys. Rev. E **60**, 1390 (1999); P. Cizeau *et al.*, Physica A **245**, 441 (1997).
[14] A. Abhayankar *et al.*, J. Bus. Finance Account., **24**, 343 (1997) documents a study of intraday patterns in bid-ask spread for the quote-driven LSE SEAQ.
[15] T. McInish and R. Wood, J. Financ. **47**, 753 (1992).
[16] A. Admati and P. Pfleiderer, Rev. Financ. Stud. **1**, 723 (1988).
[17] V. Plerou *et al.*, Phys. Rev. E **62**, R3023 (2000).
[18] M. Lundin *et al.*, in *Financial Markets Tick by Tick*, edited by P. Lequeux (Wiley, New York, 1999), p. 91; Z. Ding *et al.*, J. Empirical Finance **1**, 83 (1993).
[19] The reason for the positive correlation between the bid-ask spread and volatility can be seen by the following argument

presented in Ref. [3]. Consider the price $P(t)=P^{*}+\epsilon(t)S(t)/2$, where P^{*} denotes a "fundamental" value around which prices move and $\epsilon(t)$ denotes *i.i.d.* variables $\in\{-1,1\}$, which implies that the variance of price changes is $\mathrm{Var}\Delta P(t)=s^2/2$, making volatility linearly dependent on S, where $\Delta P(t)\equiv P(t+1)-P(t)$.

[20] P. Gopikrishnan *et al.*, Phys. Rev. E **62**, 4493 (2000).

[21] Fitting a power-law function gives generally worse quality fits. The exponents that we obtain thus are quite small $\approx 0.1-0.2$, which is consistent with a logarithmic relation. We find similar results for other logarithmic relationships reported herein.

[22] J.-P. Bouchaud, M. Mezard, and M. Potters, eprint cond-mat/0203511.

[23] M. Potters and J.-P. Bouchaud, eprint cond-mat/0210710.

[24] V. Plerou *et al.*, Phys. Rev. E **66**, 027104 (2002).

[25] We took the difference between the number of buyer-initiated and seller-initiated trades into consideration because price changes, when conditioned on the number of trades, do not show a significant dependence on volume per trade [C. Jones, G. Kaul, and M. Lipson, Rev. Financ. Stud. **7**, 631 (1994)].

[26] D. Challet and R. Stinchcombe, eprint cond-mat/0211082; eprint cond-mat/0208025. See also Physica A **300**, 285 (2001).

[27] G. Zumbach, eprint cond-mat/0407769, Quantitative Finance (to be published).

[28] S. Maslov and M. Mills, Physica A **299**, 234 (2001).

[29] E. Smith *et al.*, Quant. Finance **3**, 481 (2003).

[30] A. G. Zawadowski, J. Kertész, and G. Andor, e-print cond-mat/0401055, Physica A (to be published); see also eprint cond-mat/0406696.