Persistent Intraday Correlations Create Skews in Daily-Scale Distributions

T. J. Mazurek

t_j_mazurek@yahoo.com

July 7, 2017

Abstract. This study shows that the known extreme persistence of sign-correlations in intraday order-flow creates a positively skewed daily-scale distribution of bull-bear sentiment. The latter is defined by the difference of daily buyer minus seller trade totals. This study then introduces the intraday price-mobility and shows that it also exhibits extreme persistence in sign-correlations of even greater strength than those of order-flow. It then shows that the correlations of price-mobility also create a positively skewed daily-scale price-change distribution. This work also introduces the normal inverse Gaussian representation (NIG) of empirical price-change and bull-bear sentiment distributions to examine their behaviors with increasing scales. It shows that the shape and skew parameters of both distributions initially have parallel behaviors with increasing scales that sharply contrast with the behaviors of NIG independent identical distributions (NIG IIDs): the skews grow instead of decaying while the shape parameters decay much more slowly than those of NIG IIDs. The extremely persistent intraday correlations create these non-IID scaling behaviors.

1 Introduction

The extreme persistence of order-flow correlations was first pointed out independently by Bouchaud, Gefen, Potters, and Wyart (2004), and Lillo and Farmer (2004). Bouchaud, Farmer, and Lillo (2009) present a review of such long memory processes. These auto correlations characterize the tendency of buy (sell) orders to be followed more buy (sell) orders. They are termed as "sign-correlations" due to the historical practice of denoting buyer orders by $+1$ and seller ones by -1. These correlations are extremely persistent in the sense that the auto correlations are positive and statistically significant out to lags of thousands of trades corresponding to trading times that span over several days.

This study differs from the ones above in that it actually determines the individual correlations for a buyer (seller) trade to be followed by more buyer (seller) ones. In this study's formulation the auto correlations used by the prior works are the average of the individual buyer and seller correlations determined here. With these individual buyer and seller auto correlations, this study can determine the probabilities of occurrence of buyer or seller trade at any specified number of lags. The results show that the difference in probabilities of buyer minus seller trades is positive at all lags. Thus the occurrence of trades is skewed in favor of those by buyers. This study derives implications of this buyer bias for the daily sentiment distribution's skew.

Let's now define the daily bull-bear sentiment specifically. The daily bull-bear sentiment is the difference of the buyer minus seller daily trade totals. The probability bias toward buyer trades in the persistent sign-correlations of the above paragraph assures that the daily buyer total will on average exceed the one of sellers. One thus expects that the daily bull-bear sentiment distribution will have a positive skew.

Let's now introduce the intraday price-mobility. The latter is defined as the sign of the difference between an intraday trade's price and the prior day's closing price. This study shows that the price-mobility also has extremely persistent correlations that are even stronger than those of order-flow. The existence of such correlations can be understood as follows. Toth, Palit, Lillo, and Farmer (2015) show that order splitting drives the long-term memory exhibited by order-flow for trading times up to a few hours. Order-splitting also is likely to be the source of long-term memory in intraday price-mobility. The intent of order-splitting is to limit the price impact of large orders and suppress big price moves. To the extent it's successful, order-splitting anchors the intra-day price-mobility to the prior day's closing price, creating the observed extremely long-term intraday sign-correlations.

The price-mobility's individual positive and negative sign-correlations as well as the average of the two are statistically significant out to lags of thousands of trades. The individual sign-correlations persist out to trading times of several days, while their average is statistically significant out to one or two days. This work determines the individual probabilities for occurrence of a prices rise and a fall from the individual positive and negative price-mobility

correlations. The results show that the difference between the probabilities of price rise minus fall is positive for all lags. Thus the direction of price-change is skewed in favor of rising prices. Let's now look at the implications of this bias on the daily scale.

The probability bias toward rising prices in the above paragraph assures that the daily closing price is also biased toward rising prices. One thus expects that the daily price-change distribution will have a positive skew once the data base of closing prices is sufficiently large to resolve the small bias in favor of positive price-change. This introduction demonstrates how this comes about.

The following paragraphs begin by describing in broader terms this study's examination of the skew of the daily price-change distribution and its relation to price-mobility using daily closing price data[1] of S&P 500 Index. Noting that the presence of significant positive skew in this index's distribution at the monthly scale is well known, it first defines skew excesses above given threshold magnitudes and then outlines their behaviors with increasing magnitude thresholds and also with increasing scales. It then narrows its focus to consider only the all-magnitude skew with no threshold. This study then examines the S&P 500 Index's empirical all-magnitude skew excesses behaviors with increasing durations. For each of the durations the average all-magnitude skew along with its standard deviation are derived from the historical closing prices. These average empirical skews approach the price-mobility skew excess as duration increases. The empirical S&P 500 Index's skew is shown to be equal to the price-mobility's daily skew excess at the highest duration in the data set. The latter price-mobility's daily skew excess is determined by the extremely persistent correlations at lags corresponding to the daily scale. This demonstrates that historical closing prices carry the imprint of price-mobility.

To consider in more detail the empirical skew characteristic of the logarithmic price-changes, one needs to define suitable skew parameters. To prepare for this definition let's first consider the positive and negative branches of the cumulative distribution for price-change[2]. For a given scale s, the positive and negative branches of the cumulative distribution at or above a given magnitude m_s are given by

$$F_{\pm,s}^{\Delta lnP^c}\left(m_s/\sigma_s^{\Delta lnP^c}\right) \;=\; \frac{1}{N_s^{tot}} \sum_{i=1}^{N_s^{tot}} \Theta\left(\pm z_{i,s}^{\Delta lnP^c}/\sigma_s^{\Delta lnP^c} - m_s/\sigma_s^{\Delta lnP^c}\right), \qquad (1.1)$$

where $z_s^{\Delta lnP^c}$ is the drift-free price-change, $\sigma_s^{\Delta lnP^c}$ the volatility, N_s^{tot} is the number of price-changes at scale s, and $\Theta(y)$ is the unit step function that equals zero for $y < 0$ and unity

[1] The S&P 500 Index's data that is used throughout this entire introduction spans the duration from 1/3/1950 to 2/3/2010 and closing prices are from Yahoo! Finance.

[2] The drift-free price-changes on the daily scale are given by $z_i^{\Delta lnP^c} = ln(P_i^c/P_{i-1}^c) - \langle ln(P_i^c/P_{i-1}^c)\rangle$, where the closing price on day i is P_i^c, the prior day's closing price is P_{i-1}^c, and $\langle ln(P_i^c/P_{i-1}^c)\rangle$ is the mean (or drift) of $ln(P_i^c/P_{i-1}^c)$ over all days in the data set. The price-changes on higher scales are detailed in Subsection 2.6.

otherwise. Now define the price-change skew excess[3] above magnitude m_s using the cumulative distributions at or above this given magnitude as

$$\varepsilon_S^{\Delta lnP^c}(m_s/\sigma_s) = \frac{F_{+,S}^{\Delta lnP^c}\left(m_s/\sigma_S^{\Delta lnP^c}\right) - F_{-,S}^{\Delta lnP^c}\left(m_s/\sigma_S^{\Delta lnP^c}\right)}{F_{+,S}^{\Delta lnP^c}\left(m_s/\sigma_S^{\Delta lnP^c}\right)+F_{-,S}^{\Delta lnP^c}\left(m_s/\sigma_S^{\Delta lnP^c}\right)} . \quad (1.2)$$

The positive skew of the empirical distribution of the S&P 500 Index at the monthly scale is well known[4]. Perhaps the clearest way to exhibit this empirical skew is to consider variations of the positive and negative branches of the cumulative distributions given in equations (1.1). These positive and negative branches of the cumulative distribution are shown in Figure 1 at the 25 and 1 day scales.

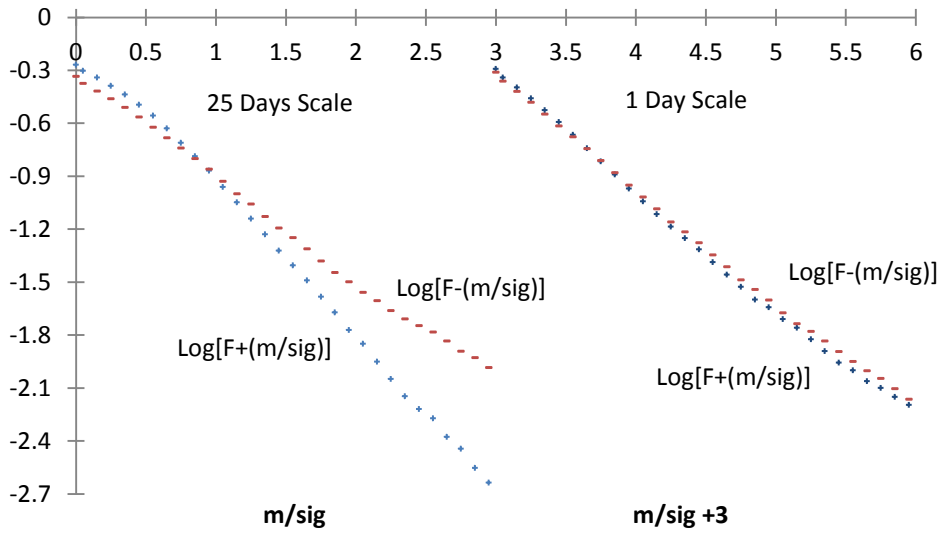


Figure 1.1 Positive and negative branches of the empirical cumulative distribution of the S&P 500 Index's closing price-changes, $F_{+,S}^{\Delta lnP^c}\left(m_s/\sigma_S^{\Delta lnP^c}\right)$ and $F_{-,S}^{\Delta lnP^c}\left(m_s/\sigma_S^{\Delta lnP^c}\right)$, respectively, on the 25 days and daily scales.

For the symmetric case, the tracks of the positive and negative branches would overlap each other. The positive[5] skew of the S&P 500 distribution on the 25 day (very roughly the monthly scale) is strikingly obvious, as the tracks of the positive and negative branches diverge significantly. At the higher magnitudes above about $0.85\sigma_{25}^{\Delta lnP^c}$ the negative branch of the

[3] The corresponding functions for the bull-bear sentiment variable are obtained via the substitutions: $\Delta lnP^c \rightarrow N^B - N^S$ and $\sigma_S^{\Delta lnP^c} \rightarrow \sigma_S^{N^B - N^S}$ in equations (1.1) and (1.2) where N^B (N^S) denote buyer (seller) trade totals.

[4] See here, or google the phrase "S&P 500 Index skew."

[5] The overall skew is positive since the total of positive price-changes exceeds that of negative ones, i.e., the skew when all magnitudes are included for $m_s = 0$ is positive: $\varepsilon_S^{\Delta lnP^c}(0) > 0$.

distribution lies above that of the positive one. The distance between them grows with increasing magnitudes so that for the highest price-change magnitude at $\sim 3\sigma_{25}^{\Delta lnP^c}$, the probability of a negative price-change is about a factor of 4.5 above that of a positive one. At low fluctuation magnitudes below about $0.85\sigma_{25}^{\Delta lnP^c}$ the positive magnitude is larger, and becomes more dominant with decreasing threshold magnitudes. The distribution on the daily scale is shown displaced by three units right-word in the figure. It also displays a positive skew, but one of an appreciably lower strength. Here the negative arm's dominance lies above magnitudes of about $0.75\sigma_{1}^{\Delta lnP^c}$, while the positive arm dominates below the latter magnitude. This figure demonstrates that the skew of the price-change distribution grows significantly from the daily to roughly the monthly scales.

Now let's confine our focus to only the all-magnitude skew excess and its relation to the price-mobility skew excess. Section 3 determines the latter to be $\varepsilon_{daily}^{PM} = 0.046(1 \pm 0.17)$. Now consider how this this skew excess would appear in daily empirical data sets with varying durations N_{tot}^d. For a given duration N_{tot}^d, the all-magnitude skew excess is $\varepsilon_{All,N_{tot}^d}^{\Delta lnP^c} = \frac{N_+ - N_-}{N_{tot}^d}$ where $N_{tot}^d = N_+ + N_-$, and the standard deviation is $\frac{1}{\sqrt{4N_{tot}^d}}$. The empirical all-magnitude price-change skew excess $\varepsilon_{All,\,N_{tot}^d}^{\Delta lnP^c}$ for N_{tot}^d has two different standard-deviation contributions, one corresponding to the intraday sample size and the other to the daily closing prices sample size. It is given

$$\varepsilon_{All,\,N_{tot}^d}^{\Delta lnP^c} = \varepsilon_{daily}^{PM}\left[1 \pm \frac{1}{\sqrt{4N_{tot}^d}}\right] = 0.046\left[1 \pm 0.17 \pm \frac{11}{\sqrt{N_{tot}^d}}\right]. \tag{1.3}$$

So that $\varepsilon_{All,\,N_{tot}^d}^{\Delta lnP^c}$ will be more consistently positive as N_{tot}^d continues to increase above the threshold given by $\frac{11}{\sqrt{N_{tot}^d}} - 0.17 = 1 \rightarrow N_{tot}^d \cong 88$.

Let's now considers behaviors of $\varepsilon_{All,\,N_{tot}}^{\Delta lnP^c}$ with increasing durations using the S&P500 Index's data set from 1/3/1950 to 2/3/2010 with total duration $N_{tot} = 15{,}115$. A given duration d contains N_{tot}^d days. One determines all[6] skew excesses of this duration $\varepsilon_d^{\Delta lnP^c}$ contained within the entire data set and computes their mean. Each period's sample size N_{tot}^d determines its standard deviation $\frac{1}{\sqrt{4N_{tot}^d}}$. Figure 1.2 below gives the results of this procedure.

[6] For duration of n years one steps through the data set in one year intervals to obtain different periods whose total is $\left\lfloor \frac{N_{tot}}{250} \right\rfloor - n + 1$, where 250 is the approximate number of trading days in a year and $\lfloor \cdot \rfloor$ denotes the greatest integer function.

One sees that at ~7 years the one-standard deviation range becomes positive. This range narrows significantly with increasing duration and at 60 years it is given by equation (1.3) to be $0.026 \pm 0.008 \pm 0.004 \leq 0.038$. This agrees at the one-standard deviation level with Section 3's range of the price-mobility excess $\varepsilon_{daily}^{PM} = 0.046 \pm 0.008 \geq 0.038$. This shows that the price-mobility correlations create the positive skew in the daily-scale distribution, and the historical closing prices carry its imprint.

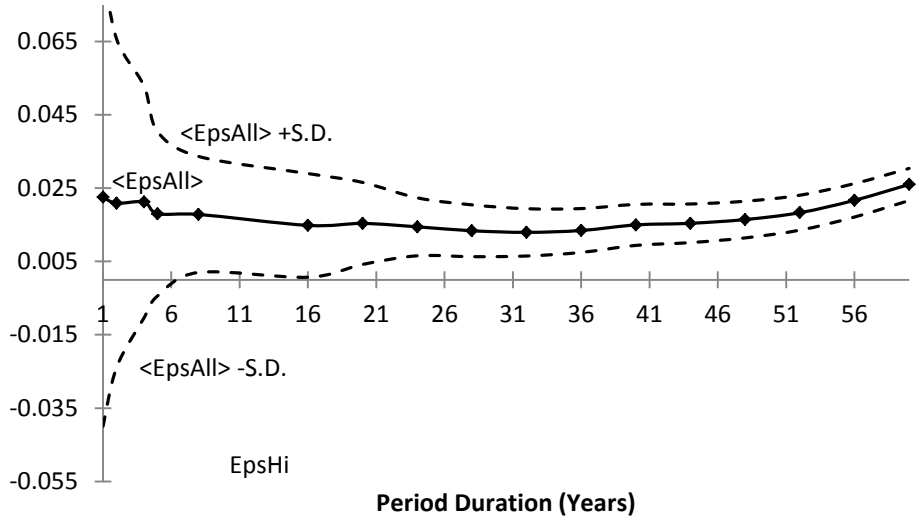


Figure 1.2 Behaviors of $\varepsilon_{All,N_{tot}^{d}}^{\Delta lnP^{c}}$, with increasing durations N_{tot}^{d} in years and the one standard deviation range.

Let's next consider the behaviors of the price-change and sentiment variable distributions with increasing scales. Gopikrishnan, Plerou, Amaral, Meyer, and Stanley (1999) present a classic study on the scaling of empirical distributions. Their primary focus was to establish that the empirical distributions had finite second moments while also exhibiting heavy power-law tails. The current work differs from theirs in that it aims to specify the empirical distributions of bull-bear sentiment and price-change at daily and higher scales completely. It does this by introducing the normal inverse Gaussian[7] (NIG) representation of empirical distributions. This representation can account for the wild behavior of Levy-like distributions and it convergence to a Gaussian distribution at high scales (see Subsection 2.9). In the NIG representation any empirical distribution is uniquely specified by its summary parameters: the mean or drift, the volatility, the all-magnitude skew excess, and the shape parameter defined by the ratio of volatility to mean absolute deviation. One can thus examine the behavior of empirical distributions with increasing scales by looking at the behavior of their summary parameters.

[7] cf. Barndorff-Nielsen (1998).

This study examines the behaviors with increasing scale of each summary parameter for price-change and bull-bear sentiment and compares them to each other. The daily skews of price-change and bull-bear sentiment should behave similarly since the order-flow and the price-mobility daily skews have the same signs and also have roughly comparable magnitudes. Also one expects bull-bear sentiment to influence price-changes and vice versa thereby promoting similar behaviors of the two. Thus the other distribution summary parameters should also have parallel behaviors with increasing scales.

This study's results show that the variances of both distributions have similar growths with increasing scales, and the growths correspond to that of identical independent distributions (IID). Both the price-change and sentiment variables have negligible daily-scale auto correlations that assure such IID variance behaviors. The means of these two variables also behave similarly with increasing scales, again exhibiting IID behaviors. The scaling behaviors of the all-magnitude skew excesses and shape parameters are more interesting.

Both price-change and sentiment all-magnitude skews grow with increasing scales. Such growths are contrary-to-IID behavior: the central limit theorem mandates that IID distributions evolve toward the Gaussian, which is symmetric, so IID skews must decrease with increasing scales. The normal-inverse-Gaussian independent identical distributions (NIG IIDs) with daily values equal to the empirical ones both decay to zero by about the seven day scale. The growths of the price-change all-magnitude skews trend steadily higher from the daily through the 25 day scale. The growth of the sentiment skew is weaker, but it does trend higher from the daily through the 15 day scales. Above the 15 day scale the sentiment skew collapses nearly to zero by the 25 day scale, with a trend that parallels the expected IID decrease with rising scales.

The initial weaker growth and subsequent collapse of the sentiment all-magnitude skew is due to noise or errors introduced by the Lee and Ready classification procedure (1991) used here to identify buyer and seller trades. This noise steadily builds with increasing scales and at scales above 16 days produces a steady collapse of the all-magnitude skew excess to values below its daily scale one. One surmises that the sentiment skew behavior would track that of the price-change skew much more closely in the absence of the noise due to classification errors.

The shape parameters of both the price-change and sentiment variables have pronounced and parallel deviations from IID behavior with increasing scales. For IIDs the central limit theorem's mandated evolution toward a Gaussian implies that the shape parameter must decrease with increasing scales and approach the Gaussian's limiting shape parameter of $\sqrt{\pi/2}$. Here the deviation from NIG IID is in the rate-of-change toward the Gaussian value. The shape parameters for the NIG IIDs with initial daily summary parameters equal to the empirical ones decay to the Gaussian limit by a scale of around six days. The empirical sentiment shape parameter starts from a value above that of the price-change one, but it decays more quickly so that it reaches the latter's value by about the seven day scale. Above the seven day scale both empirical price-change and sentiment shape parameters plateau at a value significantly greater

than the Gaussian limit out to the 15 day scale. This is almost three times the range of the NIG IID decay. The price-change shape parameter maintains its plateau out to the 25 day scale. The sentiment's scale parameter decays steadily beyond the 15 day scale and approaches the Gaussian limit at around the 25 day scale. This decay in the sentiment shape parameter is again due to the errors introduced by use of the Lee and Ready classification of buyer and seller trades.

The remainder of this paper is structured as follows. Section 2 describes the historical trade and quote data set used in this study for iShares S&P 500 Value ETF. It also presents detailed summaries of all analysis methods employed in this study. Section 3 details the intraday sign-correlations results for order-flow and price-mobility. Section 4 describes the behaviors of the price-change and sentiment variable distributions with increasing scales. The detailed behavior each distribution's individual summary parameters with increasing scales are presented. The corresponding parameter behaviors for price-change and sentiment are compared to each other. Section 5 presents behaviors with increasing lag times of daily-scale auto correlations of the buyer and seller daily trade totals. It summarizes the correlation characteristics and considers what could determine their shape of decay with increasing lags. Section 6 summarizes all of this study's results. It outlines how the positive biases created by the extremely persistent sign-correlations at the daily-scale induce skew growths with increasing scales. This section ends with conjectures on how trader activities could give rise to the negative skews typically observed at high magnitudes of the distributions above around 1.4 times the volatility.

Three appendices follow the last section. Appendix 1 presents relations between higher-than-daily scale variances and daily auto correlations. Appendix 2 gives rough estimates of noise due to errors introduced by the Lee and Ready classification method at the daily and higher scales. Appendix 3 presents daily volume-weighed totals' auto correlations and argues that their characteristics as well as the characteristics of the correlations in Section 5 derive from the period of the IVE data set which contains the great recession.

2. The S&P 500 Value ETF Data Set and Analysis Methods

This section describes the historical trade and quote data set used in this study for iShares S&P 500 Value ETF, and presents detailed summaries of all analysis methods employed in this study.

Subsection 2.1 describes the trade-and-quote data set. Subsection 2.2 outlines the Lee-Ready Classification of buyer and seller trades. Subsection 2.3 defines the Lee-Ready daily trade totals and volume-weigh totals. Subsection 2.4 presents the relations between the true buyer (seller) totals and volumes as well as the true sentiment to their respective Lee-Ready values. It also describes how the true probabilities are related to the Lee-Ready ones and summarizes empirical values of Lee-Ready errors. Subsections 2.5 and 2.6 define the daily and higher scaled variables:

the logarithmic price-change, the buyer (seller) trade total and volume, and the bull-bear sentiment as measured by the difference of buyer minus seller trade totals. Subsection 2.7 details the intraday sign-correlation equations of order-flow which determine the probabilities of occurrence of buyer and seller trades at any given lag total. It also presents the corresponding quantities for the intraday sign-correlation of price-mobility. Subsection 2.8 describes cross and auto correlations. Subsection 2.9 introduces the Normal Inverse Gaussian representations of empirical distributions and summarizes this distribution's independent identical scaling behavior. Subsection 2.10 defines the daily average buyer (seller) price-changes per tick and shows their distribution; averages of the latter quantities over all days in the data set are then used to express the sentiment variable in units of price-change.

2.1 The Trade and Quote Data Set

The trade and quote data set was downloaded from Kibot on September 8, 2014. This site offers free trade and quote data for the exchange traded fund iShares S&P 500 Value ETF with ticker symbol IVE. The data set spans about five years of trading days, beginning September 28, 2009 and ending September 8, 2014. Hereafter this data set is referred to by only its ticker symbol as the IVE data set. For each sequential trade occurrence or tick within it the data set lists the following:

		Trade	Price	Price	Trade
Date	*Time*	*Price*	*Bid*	*Ask*	*Size*
Month/Day/Year	Hour:Minute:Second	$	$	$	#Shares

Note that the prices may take fractional cent values.

2.2 Lee-Ready Classification of Buyer and Seller Trades

To classify buyer (seller) trades this study uses a method due to Lee and Ready (1991). The specific form here is as follows. From the bid and ask prices of each trade one calculates the mid-spread price by taking their average. Then if the trade's price is above (below) the mid-spread price the trade is assumed to be buyer (seller) initiated. If the trade's price equals the mid-spread price then the trade is classified to be the same as the preceding one. This uniquely defines all of the daily tick trades. Note that for all daily opening trades the prior day's closing trade's type is the preceding one. This may not be appropriate for every daily opening trade

since price changing news can occur over night. However, the averages of the daily trade totals are each around 1000. Thus sporadic incorrect assignments of opening trades should introduce very small errors to any one buyer or seller daily trade total (typically below around 0.1%). Since the preceding trade type determines the current one's type only if the opening price exactly equals the mid-spread price, the errors introduced by specifically this procedure are negligible.

2.3 The Lee-Ready Daily Trades Totals and Volumes

This subsection specifies the buyer (seller) initiated Lee-Ready daily trade total $N_i^{B,LR}$ ($N_i^{S,LR}$) and volume $V_i^{B,LR}$ ($V_i^{S,LR}$).

These are derived from the individual intraday tick trades as follows:

$$N_i^{B,LR} = \sum_{t=t_i^{open}}^{t=t_i^{close}} \sum_{i_t=0}^{i_t=n_{t_i}^{max}} \Theta^{LRM}[B] i_t \text{ and } N_i^{S,LR} = \sum_{t=t_i^{open}}^{t=t_i^{close}} \sum_{i_t=0}^{i_t=n_{t_i}^{max}} \Theta^{LRM}[S] i_t \,. \quad (2.3.1)$$

Here i_t denotes the tick trades at a given hour, minute, and second (or time) on the date of day i, with t_i^{open} and t_i^{close} being the opening and closing tick trades of that day, respectively. The total number of tick trades at any given intraday time can vary from zero to some variable maximum $n_{t_i}^{max}$. Finally, $\Theta^{LRM}[B]$ ($\Theta^{LRM}[S]$) equals unity for buyer (seller) trades and is zero for seller (buyer) initiated ones. The Lee and Ready method (LRM) described in the prior subsection defines each of these functions.

The daily buyer and seller initiated trades can be weighted by each trade's volume V_{i_t} to give the Lee-Ready daily volumes of buyer and seller initiated trades

$$V_i^{B,LR} = \sum_{t=t_i^{open}}^{t=t_i^{close}} \sum_{i_t=0}^{i_t=n_{t_i}^{max}} V_{i_t} \Theta^{LRM}[B] i_t \text{ and } V_i^{S,LR} = \sum_{t=t_i^{open}}^{t=t_i^{close}} \sum_{i_t=0}^{i_t=n_{t_i}^{max}} V_{i_t} \Theta^{LRM}[S] i_t \,, \quad (2.3.2)$$

where $V_i^{B,LR}$ and $V_i^{S,LR}$ are the total daily volumes of buyer and seller initiated trades, respectively. One has $V_i^{B,LR} + V_i^{S,LR} = V_i^T$ where V_i^T is the total true share-volume on day i.

2.4 The True Quantities in Terms of to the Lee-Ready Ones

This subsection determines the relations between the daily true trade totals, volumes, and sentiment in terms of the daily Lee-Ready ones. It also gives the true probabilities for buyer and

seller trades in terms of the Lee-Ready ones. It ends after summarizing empirical estimates of all the Lee-Ready errors.

Denote the buyer and seller classification errors by e_B^{LR} and e_S^{LR}, respectively. Specifically $e_{B\,(S)}^{LR}$ is the fraction of all true buyer (seller) trades that are erroneously classified as seller (buyer) ones. Then the Lee and Ready total of buyer (seller) trades $N_i^{B(S),LR}$ are given in terms of the true buyer (seller) totals $N_i^{B(S)}$ by

$$N_i^{B,LR} = (1 - e_B^{LR})N_i^B + e_S^{LR}N_i^S = N_i^B - e_B^{LR}N_i^B + e_S^{LR}N_i^S \quad \text{and}$$

$$N_i^{S,LR} = (1 - e_S^{LR})N_i^S + e_B^{LR}N_i^S = N_i^S - e_S^{LR}N_i^S + e_B^{LR}N_i^B \; . \tag{2.4.1}$$

One obviously has the identity

$$N_i^{B,LR} + N_i^{S,LR} = N_i^B + N_i^S \; . \tag{2.4.2}$$

Let's now express equations (2.4.1) in terms of the symmetric and asymmetric Lee-Ready errors $e_{sym}^{LR} = (e_B^{LR} + e_S^{LR})/2$ and $e_{asym}^{LR} = (e_B^{LR} - e_S^{LR})$, respectively. With $e_B^{LR} = e_{sym}^{LR} + e_{asym}^{LR}/2$ and $e_S^{LR} = e_{sym}^{LR} - e_{asym}^{LR}/2$ one gets

$$N_i^{B,LR} = N_i^B - e_{sym}^{LR}\left(N_i^B - N_i^S\right) - \frac{e_{asym}^{LR}}{2}\left(N_i^B + N_i^S\right) \quad \text{and}$$

$$N_i^{S,LR} = N_i^S + e_{sym}^{LR}\left(N_i^B - N_i^S\right) + \frac{e_{asym}^{LR}}{2}\left(N_i^B + N_i^S\right) \tag{2.4.3}$$

Solving the above equations for the true values one obtains

$$N_i^B = \frac{1}{2\left(1-2e_{sym}^{LR}\right)}\left(N_i^{B,LR} - N_i^{S,LR}\right) + \left(\frac{1}{2} + \frac{e_{asym}^{LR}}{2\left(1-2e_{sym}^{LR}\right)}\right)\left(N_i^{B,LR} + N_i^{S,LR}\right) \quad \text{and}$$

$$N_i^S = -\frac{1}{2\left(1-2e_{sym}^{LR}\right)}\left(N_i^{B,LR} - N_i^{S,LR}\right) + \left(\frac{1}{2} - \frac{e_{asym}^{LR}}{2\left(1-2e_{sym}^{LR}\right)}\right)\left(N_i^{B,LR} + N_i^{S,LR}\right). \tag{2.4.4}$$

Thus the difference of the true totals in terms of the Lee and Ready totals is

$$N_i^B - N_i^S = \frac{1}{\left(1-2e_{sym}^{LR}\right)}\left(N_i^{B,LR} - N_i^{S,LR}\right) + \frac{e_{asym}^{LR}}{\left(1-2e_{sym}^{LR}\right)}\left(N_i^{B,LR} + N_i^{S,LR}\right). \tag{2.4.5}$$

The true volume relations are obtained by the substitution $N \to V$.

Dividing on the left of the equal sign in each equation (2.4.3) by $\left(N_i^{B,LR} + N_i^{S,LR}\right)$ and the right by $\left(N_i^B + N_i^S\right)$ one gets the probabilities of having a Lee and Ready buy (sell) in terms of the terms of the true buy (sell)

$$P_B^{LR} = P_B - e_{sym}^{LR}(P_B - P_S) - \frac{e_{asym}^{LR}}{2} \quad \text{and} \quad P_S^{LR} = P_S + e_{sym}^{LR}(P_B - P_S) + \frac{e_{asym}^{LR}}{2}. \qquad (2.4.6)$$

Solving equations (2.4.6) for the true probabilities one obtains

$$P_B = \frac{1}{1-2e_{sym}^{LR}}\left(P_B^{LR} - e_{sym}^{LR} + \frac{e_{asym}^{LR}}{2}\right) \quad \text{and} \quad P_S = \frac{1}{1-2e_{sym}^{LR}}\left(P_S^{LR} - e_{sym}^{LR} - \frac{e_{asym}^{LR}}{2}\right). \qquad (2.4.7)$$

Finally we define the Lee-Ready probabilities P_B^{LR} and P_S^{LR} in terms of the sign-correlations for order-flow. From Subsection below 2.7 one has the relations $P_{BB}^{LR} = \frac{1}{4}(1 + C_{BB}^{LR})$ and $P_{SS}^{LR} = \frac{1}{4}(1 + C_{SS}^{LR})$ where the lag-time designations are suppressed here. The former probabilities are

$$P_B^{LR} = \frac{P_{BB}^{LR}}{P_{BB}^{LR}+P_{SS}^{LR}} = \frac{(1+C_{BB}^{LR})}{2+(C_{BB}^{LR}+C_{SS}^{LR})} \quad \text{and} \quad P_S^{LR} = \frac{P_{SS}^{LR}}{P_{BB}^{LR}+P_{SS}^{LR}} = \frac{(1+C_{SS}^{LR})}{2+(C_{BB}^{LR}+C_{SS}^{LR})}. \qquad (2.4.8)$$

Equations (2.4.7) then provide the true probabilities.

Odders-White (2000) provides empirical data on classification errors by the Lee-Ready method. Table 2.4.1 reproduces the relevant data for determining these errors from her Table 2. The table also gives all of the resulting Lee-Ready errors and their sample-size standard deviations (S.D.).

Table 2.4.1 Lee-Ready errors and standard deviations (S.D.) determined from empirical data by Odders-White, 2000.

Panel C: Lee and Ready method vs. true classification			
Classified by:	True Buy	True Sell	
Lee and Ready method as Buy	144348	24183	
Lee and Ready method as Sell	23680	126153	
Totals	168028	150336	
	e_B^{LR}	e_S^{LR}	S.D.
Misclassified fraction	0.140928893		0.009063
Misclassified fraction		0.16086	0.009314
$e_{asym}^{LR} = e_B^{LR} - e_S^{LR} =$		-0.01993	0.018376
$e_{sym}^{LR} = (e_B^{LR} + e_S^{LR})/2 =$		0.150894	0.018376

2.5 Summary Parameters of Daily Distributions

This subsection defines the summary parameters that describe daily distributions of buyer (seller) trade totals, volumes, bull-bear sentiment, and logarithmic closing prices.

Let Y_i^X represent the daily variable on day i, where X denotes one of the following: buyer (seller) daily total $N^B(N^S)$, buyer (seller) daily volume $V^B(V^S)$, bull-bear sentiment $N^B - N^S$, or logarithmic closing price-change ΔlnP^c. The mean or drift of Y_i^X over the total of days N_{tot} in the data set and its drift-free variable are respectively

$$\langle Y_i^X \rangle = \frac{1}{N_{tot}} \sum_{i=1}^{i=N_{tot}} Y_i^X \quad \text{and} \quad Z_i^X = Y_i^X - \langle Y_i^X \rangle. \tag{2.5.1}$$

The volatility σ^X and mean absolute deviation $\tilde{\sigma}^X$ of Y_i^X are respectively given by

$$(\sigma^X)^2 = \frac{1}{N_{tot}} \sum_{i=1}^{i=N_{tot}} (Z_i^X)^2 \quad \text{and} \quad \tilde{\sigma}^X = \frac{1}{N_{tot}} \sum_{i=1}^{i=N_{tot}} \left| Z_i^X \right|. \tag{2.5.2}$$

For distributions having negative and positive ranges one defines two additional parameters: one for shape and one for skew. The distribution's shape parameter is given by the ratio of volatility to mean absolute deviation $\sigma^X/\tilde{\sigma}^X$. The distribution's skew is characterized by the ratio of the total positive to total negative drift-free fluctuations R_{pn}^X. The all-magnitude skew excess discussed in Section 1 is an equivalent antisymmetric skew characterization, and the relation between the two is

$$\varepsilon^X(0) = \left(R_{pn}^X - 1\right)/\left(R_{pn}^X + 1\right). \tag{2.5.3}$$

Thus the distributions of logarithmic daily closing price-changes with variables $\Delta lnP_i^c = ln(P_i^c/P_{i-1}^c)$ and of daily bull-bear sentiment with variables $(N_i^B - N_i^S)$ are defined by the summary parameters $\langle Y_i^X \rangle$, $\tilde{\sigma}^X$, σ^X, $\sigma^X/\tilde{\sigma}^X$, and $\varepsilon^X(0)$, where X represents either one. Table 2.5.1 gives the summary parameter of buyer (seller) daily trade totals along those of sentiment and price-change.

Table 2.5.1 Distribution summary parameters for daily variations in buyer (seller) trade totals, sentiments, and prices.

X	$\langle X_i \rangle$	$\tilde{\sigma}^X$	σ^X	$\sigma^X/\tilde{\sigma}^X$	$\varepsilon^X(0)$
N^B	995.427	402.586	570.012		
N^S	945.651	411.332	576.284		
$N^B - N^S$	49.776	178.786	300.235	1.679	0.039
ΔlnP^c	0.000466	0.00736	0.010558	1.435	0.053

Some actual empirical distributions of price-change and sentiment will be considered in the Subsection 2.9. The volume distributions will not be considered.

2.6 Summary Parameters of Distributions above the Daily Scale

This subsection considers explicitly only the bull-bear sentiment and price-change variables. The buyer (seller) total variable is similar to the sentiment variable and is obtained from the sentiment equations below by the substitutions $N^B - N^S \to N^B$ (N^S) while maintaining all original subscripts.

Extending this procedure of Subsection 2.4 to higher scales $s > 1$, one defines the sentiment variables at scale s as sums over a range of similar daily variables

$$\left(N^B_{j,s} - N^S_{j,s}\right) = \sum_{i=1}^{s}\left(N^B_{j-i+1,1} - N^S_{j-i+1,1}\right) \tag{2.6.1}$$

where for clarity the daily scale is now explicitly noted on the right in the subscripts after the comma. The index j satisfies $s \le j \le N_{tot}$. The mean or drift of the daily bull-bear sentiment is

$$\langle N_j^{N^B-N^S}\rangle_s \equiv \langle N^B_{j,s} - N^S_{j,s}\rangle = \frac{1}{N_{tot}-s}\sum_{j=s}^{j=N_{tot}}\left(N^B_{j,s} - N^S_{j,s}\right). \tag{2.6.2}$$

Subtracting this drift from each of the original series elements gives the drift-free series $Z^{N^B-N^S}_{j,s} = \left(N^B_{j,s} - N^S_{j,s}\right) - \langle N_j^{N^B-N^S}\rangle_s$. The ratio of the total positive to the total negative drift-free fluctuations determines $R^{N^B-N^S}_{pn,s}$. The all-magnitude skew excess is

$$\varepsilon_s^{N^B-N^S}(0) = \left(R^{N^B-N^S}_{pn,s} - 1\right)/\left(R^{N^B-N^S}_{pn,s} + 1\right). \tag{2.6.3}$$

The volatility and mean absolute deviation are given respectively by

$$\left(\sigma_s^{N^B-N^S}\right)^2 = \frac{1}{N_{tot}-s}\sum_{i=s}^{i=N_{tot}}\left(Z^{N^B-N^S}_{i,s}\right)^2 \qquad \text{and}$$

$$\tilde{\sigma}_s^{N^B-N^S} = \frac{1}{N_{tot}-s}\sum_{i=s}^{i=N_{tot}}\left|Z^{N^B-N^S}_{i,s}\right|. \tag{2.6.4}$$

Finally the shape parameter is $\sigma_s^{N^B-N^S}/\tilde{\sigma}_s^{N^B-N^S}$.

The logarithmic price-change variable at scale s is the following sum over daily closing price-changes

$$(\Delta lnP^c)_{j,s} = \sum_{i=1}^{i=s} ln\left(P^c_{j-i+1,1}/P^c_{j-i,1}\right) = ln\left(P^c_{j,1}/P^c_{j-s,1}\right), \tag{2.6.5}$$

with j again satisfying $s \le j \le N_{tot}$. Note that the higher scale price-change variables are sums over daily price-changes, similar to the sums over daily sentiment changes. However, here the intermediate terms cancel leaving only the differences between starting and ending terms.

The mean or drift $\langle (\Delta lnP^c)_{j,s}\rangle$ is obtained by averaging over j as in equation (2.6.2). We denote this mean by the symbol $\langle \Delta lnP^c\rangle_s$. The drift-free series $Z^{\Delta lnP^c}_{j,s} = (\Delta lnP^c)_{j,s} - \langle \Delta lnP^c\rangle_s$

determine the total positive to total negative fluctuation ratio $R_{pn,s}^{\Delta lnP^c}$ and the all-magnitude skew excess $\varepsilon_s^{\Delta lnP^c}(0)$. Equations corresponding to those of equations (2.6.4) determine the volatility $\sigma_s^{\Delta lnP^c}$ and the mean absolute deviation $\tilde{\sigma}_s^{\Delta lnP^c}$, as well as the price-change shape parameter $\sigma_s^{\Delta lnP^c}/\tilde{\sigma}_s^{\Delta lnP^c}$.

At each scale s the sentiment distribution's summary parameters are given by $\langle N_j^{N^B-N^S}\rangle_s$, $\sigma_s^{N^B-N^S}$, $\sigma_s^{N^B-N^S}/\tilde{\sigma}_s^{N^B-N^S}$, and $\varepsilon_s^{N^B-N^S}(0)$. The price-change distribution's ones are given by $\langle N_j^{\Delta lnP^c}\rangle_s$, $\sigma_s^{\Delta lnP^c}$, $\sigma_s^{\Delta lnP^c}/\tilde{\sigma}_s^{\Delta lnP^c}$, and $\varepsilon_s^{\Delta lnP^c}(0)$. The behaviors with increasing scales of each corresponding pair of parameters for price-change and sentiment will be presented graphically in Section 4.

2.7 Intraday Trading Order-Flow and Price-Mobility Sign Correlations

This subsection first discusses the procedure for determining intraday order-flow and price-change sign correlations. The subsection then describes the corresponding procedure for determining the intraday sign-correlation for price-mobility.

First consider intraday order-flow sign-correlations. One is interested in the tick by tick correlation of the orders, i.e., the extent to which a buy (sell) order is followed by a buy (sell). Correlations are given by the two matched combinations (B,B) and (S,S), while the mismatched combinations $(B,S),(S,B)$ give correlation disruption. Define the sign-correlation parameters as $C_{BB}(\tau)$, $C_{SS}(\tau)$, and $C_{BS}(\tau)$. One then obtains the probabilities for the matched combinations $P_{BB} = \frac{1}{4}[1 + C_{BB}(\tau)]$, $P_{SS} = \frac{1}{4}[1 + C_{SS}(\tau)]$, and the mismatched one $P_{BS} = \frac{1}{2}[1 + C_{BS}(\tau)]$. These probabilities must sum to unity so that

$$C_{BS}(\tau) = -\frac{1}{2}[C_{BB}(\tau) + C_{SS}(\tau)]. \qquad (2.7.1)$$

One then then determines these empirical correlations at intraday lag-time τ by counting the empirical total occurrences of buy-buy N_{BB}, sell-sell N_{SS}, and buy-sell N_{BS} trade pairs within the trade total N_{Tot} as follows

$$\left(\frac{N_{BB}}{N_{Tot}-\tau}\right)_\tau^{emp} = \frac{1}{4}[1 + C_{BB}(\tau)], \qquad \left(\frac{N_{SS}}{N_{Tot}-\tau}\right)_\tau^{emp} = \frac{1}{4}[1 + C_{SS}(\tau)],$$

and

$$\left(\frac{N_{BS}}{N_{Tot}-\tau}\right)_\tau^{emp} = \frac{1}{2}\left[1 - \frac{1}{2}[C_{BB}(\tau) + C_{SS}(\tau)]\right]. \qquad (2.7.2)$$

Thus the matched correlations are given by

$$C_{BB}(\tau) = 4\left(\frac{N_{BB}}{N_{Tot}-\tau}\right)_{\tau}^{emp} - 1 \quad \text{and} \quad C_{SS}(\tau) = 4\left(\frac{N_{SS}}{N_{Tot}-\tau}\right)_{\tau}^{emp} - 1. \tag{2.7.3}$$

Now adding the first two and subtracting the third of equations (2.7.2) gives

$$\bar{C}_{OF}(\tau) \equiv \frac{1}{2}[C_{BB}(\tau) + C_{SS}(\tau)] = \left(\frac{N_{BB}+N_{SS}}{N_{Tot}-\tau}\right)_{\tau}^{emp} - \left(\frac{N_{BS}}{N_{Tot}-\tau}\right)_{\tau}^{emp}, \tag{2.7.4}$$

where the left-most expression defines the average order-flow sign auto correlation $\overline{C}_{OF}(\tau)$.

Now consider the standard deviation associated with the right side of equation (2.7.4). This equation gives the average sign correlation in terms of the difference between matched and mismatched fraction totals of paired signs. This is a binary pair of possibilities where the binomial distribution applies to the right side of equation (2.7.4). The probability of matched and mismatched signs is given by p and $1-p$, respectively. The variance of the binomial distribution is $\frac{p(1-p)}{(N_{tot}-\tau)}$, and it has a maximum of $\frac{1}{4(N_{tot}-\tau)}$ for $p = \frac{1}{2}$. Thus the standard deviation for the empirical right-most expression in equation (2.7.4) is $\frac{1}{\sqrt{4(N_{tot}-\tau)}}$, and to within three standard deviations the average order-flow sign correlation is given by

$$\bar{C}_{OF}(\tau) \equiv \frac{1}{2}[C_{BB}(\tau) + C_{SS}(\tau)]\left(1 \pm \frac{3}{\sqrt{4(N_{tot}-\tau)}}\right). \tag{2.7.5}$$

Let's now determine $C_{BB}(\tau)$ and $C_{SS}(\tau)$ and their standard deviations. In equations (2.7.2) take the difference and then sum of the first two to get

$$\left(\frac{N_{BB}}{N_{Tot}-\tau}\right)_{\tau}^{emp} - \left(\frac{N_{SS}}{N_{Tot}-\tau}\right)_{\tau}^{emp} = \frac{1}{4}[C_{BB}(\tau) - C_{SS}(\tau)] \text{ and}$$

$$\left(\frac{N_{BB}}{N_{Tot}-\tau}\right)_{\tau}^{emp} + \left(\frac{N_{SS}}{N_{Tot}-\tau}\right)_{\tau}^{emp} = \frac{1}{2} + \frac{1}{4}[C_{BB}(\tau) - C_{SS}(\tau)].$$

Now divide the former equation by the latter to obtain

$$\left(\frac{N_{BB}}{N_{BB}+N_{SS}}\right)_{\tau}^{emp} - \left(\frac{N_{SS}}{N_{BB}+N_{SS}}\right)_{\tau}^{emp} = \frac{C_{BB}(\tau)-C_{SS}(\tau)}{2[1+\bar{C}_{OF}(\tau)]} \tag{2.7.6}$$

Multiplying equation (2.7.6) by the right hand side's quantity in square brackets one obtains

$$\frac{C_{BB}(\tau)}{2} - \frac{C_{SS}(\tau)}{2} = \left(1 + \bar{C}_{OF}(\tau)\right)\left[\left(\frac{N_{BB}}{N_{BB}+N_{SS}}\right)_{\tau}^{emp} - \left(\frac{N_{SS}}{N_{BB}+N_{SS}}\right)_{\tau}^{emp}\right]. \tag{2.7.7}$$

The logic of the argument in the paragraph prior to equation (2.7.5) now applies to buyer and seller correlations of equation (2.7.7), with $N_{Tot} - \tau$ being replaced by $N_{tot}^{B+S} = N_{BB} + N_{SS}$. Hence the three standard deviations here $\frac{3}{\sqrt{4N_{tot}^{B+S}}}$, apply to the term in the square brackets.

Neglecting second order terms arising from multiplication by the average sign correlation, one

thus obtains that the standard deviation for the half difference between the buy-buy and sell-sell correlations

$$\frac{C_{BB}(\tau)}{2}-\frac{C_{SS}(\tau)}{2}=\left(1+\bar{C}_{OF}(\tau)\right)\left[\left(\frac{N_{BB}}{N_{BB}+N_{SS}}\right)_{\tau}^{emp}-\left(\frac{N_{SS}}{N_{BB}+N_{SS}}\right)_{\tau}^{emp}\right]\left[1\pm\frac{3(1+\bar{C}_{OF})}{\sqrt{4N_{tot}^{B+S}}}\right]. \quad (2.7.8)$$

Equation (2.7.8) gives the difference of half the buyer and seller correlation while equation (2.7.4) gives their sum. Thus addition (subtraction) of the former half difference to the latter half sum gives $C_{BB}(\tau)$ $\left(C_{SS}(\tau)\right)$ and its three standard deviations

$$C_{BB(SS)}(\tau)=C_{BB(SS)}(\tau)\left\{1\pm\frac{3}{\sqrt{4(N_{tot}-\tau)}}\left[1+(1+\bar{C}_{OF})\sqrt{\frac{N_{tot}-\tau}{N_{tot}^{B+S}}}\right]\right\}$$

$$\cong C_{BB(SS)}(\tau)\left(1\pm\frac{3\left(1+\sqrt{2}\right)}{\sqrt{4(N_{tot}-\tau)}}\right). \quad (2.7.9)$$

The last expression in equation (2.7.9) follows for $\bar{C}_{OF}\ll 1$ and $\frac{N_{tot}-\tau}{N_{tot}^{B+S}}\cong 2$.

Now consider the corresponding tick-level sign-correlations for intraday price-mobility on day i from the prior days close P_{i-1}^{c}. The intraday price-changes are given by $P_{i_t}-P_{i-1}^{c}$ where P_{i_t} is an intraday trade's price within day i. The sign correlations are determined by the variable $\delta_{pm}=\ sign\left(P_{i_t}-P_{i-1}^{c}\right)$. The price-mobility sign correlation takes on values $\delta_{pm}\in[+1,0,-1]$ so the price-mobility has the added complication of zero values. One addresses this as follows. Temporarily ignore zero values and compute the sign correlations $\tilde{C}_{++}(\tau)$, $\tilde{C}_{--}(\tau)$, and $\tilde{C}_{+-}(\tau)$ by using the corresponding above equations for order-flow with the following substitution: $BB\rightarrow++$, $SS\rightarrow--$, and $BS\rightarrow+-$. Now take the total N_0^{tot} of $\delta_{pm}\delta_{pm}^{\tau}=0$ terms (δ_{pm}^{τ} is the lagged variable) and multiply each of the latter correlations by the factor[8] $F_0=\left(1-\frac{N_0^{tot}}{N_{Tot}-\tau}\right)$ where $N_{Tot}-\tau$ is the total number at lag τ of all correlation terms $\delta_{pm}\delta_{pm}^{\tau}$ to obtain the price-mobility correlations: $C_{++}(\tau)=F_0\tilde{C}_{++}(\tau)$, $C_{--}(\tau)=F_0\tilde{C}_{--}(\tau)$, and $\bar{C}_{PM}=-F_0\tilde{C}_{+-}(\tau)$.

The price-mobility correlations can be used to define the probabilities of occurrence of price rises and falls. One has $P_{++}(\tau)=\frac{1}{4}[1+C_{++}(\tau)]$ and $P_{--}(\tau)=\frac{1}{4}[1+C_{--}(\tau)]$. In terms of the price-mobility correlations the probabilities of price rises and falls are

$$P_{+}(\tau)=\frac{P_{++}(\tau)}{P_{++}(\tau)+P_{--}(\tau)}=\frac{1+C_{++}(\tau)}{2+C_{++}(\tau)+C_{--}(\tau)}\ \text{ and }\ P_{-}(\tau)=\frac{1+C_{--}(\tau)}{2+C_{++}(\tau)+C_{--}(\tau)}. \quad (2.7.10)$$

[8] In practice this factor is very close to unity since in the IVE data set $N_{\delta_{pm}=0}^{Tot}=17{,}650$ while $N_{\delta_{pm}}^{Tot}=2{,}416{,}641$.

Similarly the order-flow correlation define the Lee-Ready probabilities of occurrence of buyer and seller trades $P_B^{LR}(\tau)$ and $P_S^{LR}(\tau)$. These are given in Subsection 2.4 by equations (2.4.8) where the procedure for obtaining the true probabilities is also described.

2.8 Cross and Auto Correlations

The cross correlations between two given series X_i and Y_i at lag days τ is the following function of the normalized drift-free vector elements

$$C(X,Y,\tau) = \frac{1}{N_{tot}-\tau}\sum_{i=\tau}^{N_{tot}} \frac{(X_i-\langle X_i\rangle)}{\sigma^X}\frac{(Y_{i-\tau}-\langle Y_i\rangle)}{\sigma^Y}\,. \tag{2.8.1}$$

The auto correlation of a series is given by equation (2.8.1) with $Y = X$.

2.9 The Normal Inverse Gaussian: Its Representations of Empirical Distributions and Independent Identical Scaling Behaviors

This subsection first presents the Normal Inverse Gaussian (NIG) distribution. It then considers the symmetric case to motivate the use of the ratio of volatility to mean absolute deviation as a scale parameter. Numerical procedures for calculating the NIG distribution from its parameters are summarized. Then the method of obtaining NIG representations from the empirical distribution's summary parameters of volatility, skew excess, and ratio of volatility to mean absolute deviation is presented. Finally this subsection describes the distribution's parameters behaviors for independent identical NIG distributions (NIG IIDs) with increasing scales.

The NIG distribution (cf. Barndorff-Nielson, 1998) in terms of the drift-free variable z is

$$nig(\sigma,\nu,\eta,z) = \frac{\nu}{\pi\sigma\left(1-\eta^2\right)^{3/4}}\frac{K_1\left[\nu^2 g(\sigma,\nu,\eta,z)\right]}{g(\sigma,\nu,\eta,z)}\times$$

$$\exp\left\{\nu^2\sqrt{1-\eta^2}+\nu\eta\left(1-\eta^2\right)^{-3/4}\left[z/\sigma+\nu\eta\left(1-\eta^2\right)^{1/4}\right]\right\}, \tag{2.9.1}$$

where $$g(\sigma,\nu,\eta,z) = \sqrt{1+\frac{\left[z/\sigma+\nu\eta\left(1-\eta^2\right)^{1/4}\right]^2}{\nu^2\left(1-\eta^2\right)^{3/2}}},$$

σ is the volatility, $K_1(\cdot)$ is the modified Bessel function of index one, and the NIG parameters[9] ν and η determine its shape and skew, respectively. The drift-free variable is given by

$$z \;=\; x-\mu-\nu\eta\sigma\left(1-\eta^2\right)^{1/4}\,, \tag{2.9.2}$$

where μ is a constant drift and x is the variable containing drifts.

Let's now consider some salient characteristics of the symmetric NIG distribution, which takes the following form

$$nig(\sigma,\nu,z) \;=\; \frac{\nu e^{\nu^2}}{\pi\sigma}\left[1+\left[z/(\nu\sigma)\right]^2\right]^{-1/2} K_1\left\{\nu^2\left[1+\left[z/(\nu\sigma)\right]^2\right]^{1/2}\right\}. \tag{2.9.3}$$

Its limiting expression as ν approaches infinity is just the Gaussian distribution:

$$\lim_{\nu\to\infty}\left[nig(\nu,\sigma,z)\right] \;=\; \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{z^2}{2\sigma^2}\right]. \tag{2.9.4}$$

This limiting behavior also holds for the skewed NIG distribution. For ν going to zero while the product νσ is held constant, the NIG distribution approaches the Lorentz-Cauchy distribution:

$$\lim_{\nu\to 0\ and\ \nu\sigma\ fixed}\left[nig(\nu,\sigma,z)\right] \;=\; \frac{1}{\pi}\,\frac{1/(\nu\sigma)}{1+\left[z/(\nu\sigma)\right]^2}\,. \tag{2.9.5}$$

The ratio of absolute moment to volatility which determines the shape parameter ν is given by (cf. Forsberg, 2002, p. 160)

$$\frac{\sigma}{\tilde{\sigma}} \;=\; \frac{\pi}{2}\,\frac{e^{-\nu^2}}{\nu K_0(\nu^2)}\,, \tag{2.9.6}$$

where $K_0(\cdot)$ is the modified Bessel function of index zero. Let's now consider limiting behaviors of the shape parameter. As $\nu\to\infty$, $\nu K_0(\nu^2)\to\sqrt{\pi/2}e^{-\nu^2}$, so that $\sigma/\tilde{\sigma}\to\sqrt{\pi/2}$ in the Gaussian limit. For the Lorentz-Cauchy case the product $\nu K_0(\nu^2)\to\nu ln(1/\nu^2)\to 0$ as $\nu\to 0$ so that $\sigma/\tilde{\sigma}\to\infty$. Thus for the symmetric NIG distribution $\sigma/\tilde{\sigma}$ varies continuously between $\sqrt{\pi/2}$ and ∞, describing distributions' behaviors that can be as mild as a Gaussian or as wild as the Lorentz-Cauchy distribution, as well as all gradations of behaviors in between. This wide ranging ability to characterize distribution behaviors should carry over to the skewed NIG case. Thus we use the ratio of volatility to mean absolute deviation as the general shape parameter. Press, Flannery, and Teukolsky (2002) provide numerical procedures for evaluating the Bessel functions in the above equations.

[9] Note the expression in equation (2.9.1) is in the drift-free frame and the parameterization here differs from Barndorff-Nielsen's; for his parameterization use the relations: $\sigma=\sqrt{\delta/\alpha},\ \nu=\sqrt{\delta\alpha}$, and $\eta=\beta/\alpha$.

Let's now summarize the application of NIG distributions to the study of empirical ones. For given values of the parameters σ,ν, and η, one can evaluate the NIG distribution via equation (2.9.1) over the entire range of z. For clarity in the following paragraphs denote σ by σ^{nig}. Integration determines the mean absolute deviation $\tilde{\sigma}^{nig} = \int_{-\infty}^{\infty} |z| nig\left(\sigma^{nig}, \nu, \eta, z\right) dz$ and the variance $\left(\sigma^{nig}\right)^2 = \int_{-\infty}^{\infty} z^2 nig\left(\sigma^{nig}, \nu, \eta, z\right) dz$. The positive and negative fractions of the cumulative distribution, respectively, are given by $F_+^{nig}(0) = \int_0^{\infty} nig\left(\sigma^{nig}, \nu, \eta, z\right) dz$ and $F_-^{nig}(0) = \int_{-\infty}^{0} nig\left(\sigma^{nig}, \nu, \eta, z\right) dz$. The latter two define the ratio of total positive to total negative fluctuations $R_{pn}^{nig} = F_+^{nig}(0)/F_-^{nig}(0)$. The shape parameter is $\sigma^{nig}/\tilde{\sigma}^{nig}$. These integrals were evaluated via the Gauss-Laguerre quadrature procedure given in Press, Flannery, and Teukolsky (2002) using 50 quadrature points.

At a given scale, the following procedure gives the NIG representation for either a price-change or sentiment empirical distribution with parameters σ^{emp}, R_{pn}^{emp}, and $\sigma^{emp}/\tilde{\sigma}^{emp}$. For notational simplicity here the scale, price-change, and sentiment designations are suppressed. One first sets $\sigma^{nig} = \sigma^{emp}$ and then uses a two dimensional Newton-Raphson iteration procedure to determine the values of ν and η that give $R_{pn}^{nig} = R_{pn}^{emp}$ and $\sigma^{nig}/\tilde{\sigma}^{nig} = \sigma^{emp}/\tilde{\sigma}^{emp}$. Relevant derivatives are evaluated numerically. One must guess sufficiently close starting values of the latter two quantities for the Newton-Raphson iteration to converge.

Now let's discuss the scaling behavior of the NIG independent identical distribution (IID). The prior paragraph's procedure determines the daily parameters ν_1 and η_1. The NIG IID volatility, shape and skew parameters at higher scales s are $\sigma_s^{nig\ IID} = \sigma_1^{nig\ IID}\sqrt{s}$, $\nu_s^{nig\ IID} = \nu_1\sqrt{s}$, and $\eta_s^{nig\ IID} = \eta_1$, respectively. The integrals and equations in the paragraph prior to the above one are then used to determine the shape parameter and the ratio of total positive to total negative fluctuations of the NIG IID at scale s.

This subsection ends with two examples of NIG representations of empirical distributions for price-changes and for sentiments at the daily and seven-day scales. The main focus of this paper is the scaling behaviors of the summary parameters and not the actual distributions in detail. However, these two examples are presented to illustrate the usefulness of the NIG representations for studies of empirical distributions.

Figures 2.9.1 and 2.9.2 below show, respectively, the IVE empirical price-change and bull-bear sentiment distributions $f_{emp}^{\Delta lnP^c}\left(z_{emp}^{\Delta lnP^c}/\sigma_{emp}^{\Delta lnP^c}\right)$ and $f_{emp}^{N^B-N^S}\left(z_{emp}^{N^B-N^S}/\sigma_{emp}^{N^B-N^S}\right)$ on the one and seven-day scales (note that the negative of the latter distribution is shown in the figures for ease of comparison). Also shown are the corresponding NIG distribution representations (solid lines) and that of the Gaussian (dashed lines).

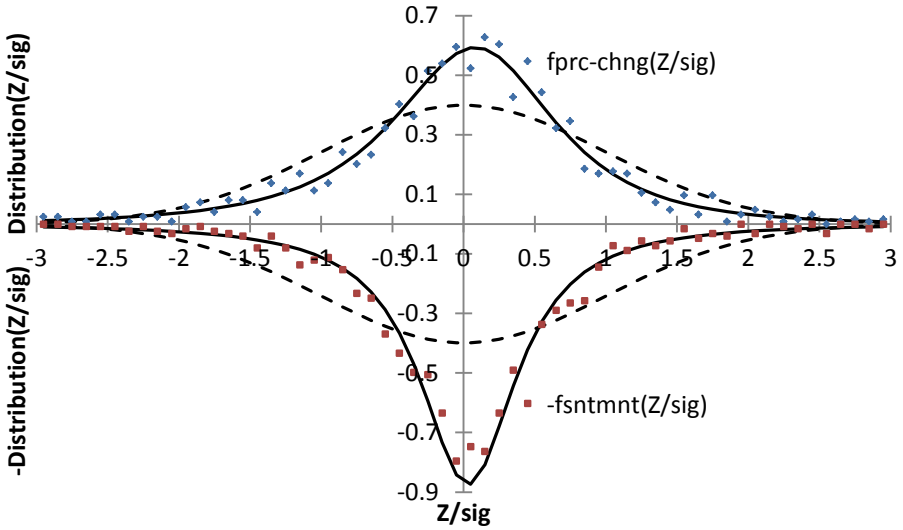


Figure 2.9.1 One-day scale IVE empirical distributions: price-change (labeled as fprc-chng) $f_{emp}^{\Delta lnP^c}\left(z_{emp}^{\Delta lnP^c}/\sigma_{emp}^{\Delta lnP^c}\right)$ and sentiment (labeled as -fsntmnt) $-f_{emp}^{N^B-N^S}\left(z_{emp}^{\Delta N^B-N^S}/\sigma_{emp}^{N^B-N^S}\right)$ (the negative of the latter distribution is shown in the figure for ease of comparison). Corresponding distributions for NIG representations (solid lines) and Gaussian (dashed lines) are also shown.

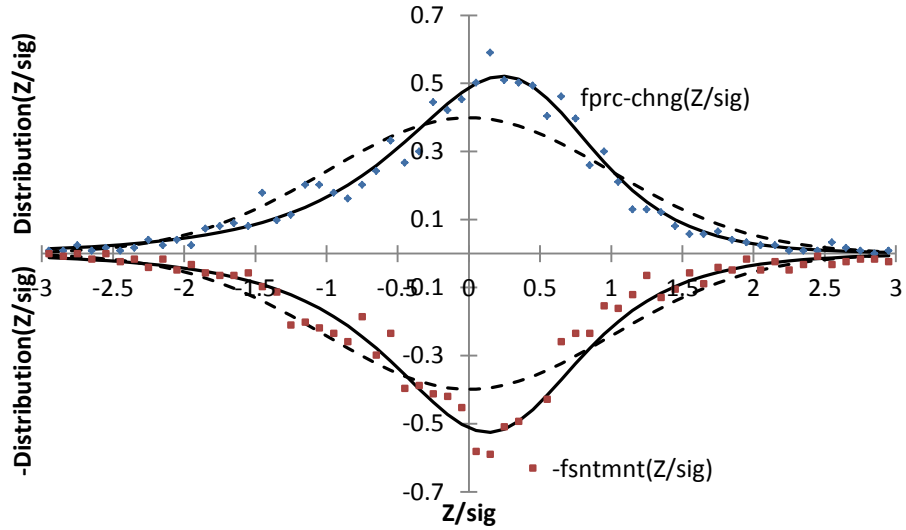


Figure 2.9.2 Seven-day scale IVE distributions. See caption of Figure 2.9.1 for details.

The daily scale figure shows that both distributions are much more strongly peaked than is the Gaussian, as is indicated by their large values of their shape parameters of 1.435 and 1.679 for price-change and sentiment, respectively. For comparison that of the Gaussian is 1.253. In turn, the sentiment distribution is much more strongly peaked than the price distribution. One can also

discern that both empirical distributions are biased toward the positive side as indicated by their positive skew excesses[10] of 0.0530 and 0.0394 for price-change and sentiment, respectively.

The seven-day-scale figure shows greater than Gaussian peaking similar to that of the daily scale, but with smaller peak values. The price-change peak is a little below its daily scale value while the sentiment peak has dropped significantly from its daily value. The shape parameters on the seven-day scale are 1.359 and 1.370 for price-change and sentiment, respectively, and have fallen below their respective daily scale values. The positive bias of both distributions has become more pronounced on the seven day scale than they were on the daily scale as is indicated by the growth of more positive skew excesses of 0.108 and 0.0670 for price-change and sentiment, respectively.

2.10 Mean Daily Buyer (Seller) Price-change per Tick and Expressing Daily Sentiment in Price-change Units

This subsection introduces daily mean buyer (seller) price-changes per tick, presents their distribution, and expresses the daily sentiment in price-change units. The daily mean buyer price changes per tick are given by

$$\langle \Delta lnP_{i_t}^{B} \rangle_i = \frac{1}{N_i^B} \sum_{t=i_{close}}^{t=i_{open}} \sum_{i_t=0}^{i_t=n_{t_i}^{max}} \left[ln\left(P_{i_t}\right) - ln\left(P_{i_t-1}\right) \right] \Theta^{LRM}[B]. \qquad (2.10.1)$$

where P_{i_t} and P_{i_t-1} denote the tick trade's price and the prior tick's one[11], respectively, and N_i^B is the daily buyer tick total. For the sellers daily price change per tick use the substitution $B \to S$.

Essentially all the daily mean price changes per tick for buyers are positive while those for sellers are negative. Of the total of 1245 mean daily price changes for each, only ten were negative for buyers and only five were positive for sellers. The flash crash price spikes on May 6, 2010 stood out[12] and were removed. The resulting average daily price changes per tick were $\langle \langle \Delta lnP_{i_t}^{B} \rangle_i \rangle = 3.17 \times 10^{-5}$ and $\langle \langle \Delta lnP_{i_t}^{S} \rangle_i \rangle = -3.25 \times 10^{-5}$.

Figure 2.10.1 below shows the IVE distributions of the buyer and seller mean daily price changes per tick. The two empirical distributions are very similar, and the figure shows in dashed lines the same generalized gamma distribution representation of each one.

[10] The skew excess is given by $\varepsilon^{emp} = (R_{pn}^{emp} - 1)/(R_{pn}^{emp} + 1)$.

[11] For each day's opening trade, P_{i_t-1} is the prior day's closing price.

[12] These showed huge daily price changes per tick of 0.0053 and -0.0038 for buyers and sellers, respectively.

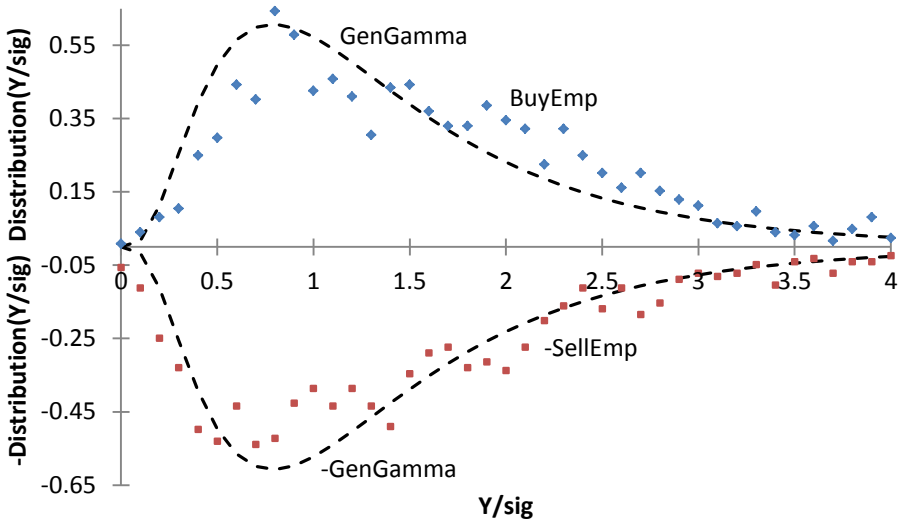


Figure 2.10.1 Buyer (seller) daily price-change per tick distribution $f^B\left(Y^B/\sigma^{\Delta lnP^B_{i_t}}\right)$ and $-f^S\left(Y^S/\sigma^{\Delta lnP^S_{i_t}}\right)$, respectively. Dashed lines are Generalized Gamma distribution[13] fits.

The daily price-change from the prior day's closing price can be written as

$$ln(P_i^c/P_{i-1}^c) = N_i^B\langle\Delta lnP_{i_t}^B\rangle_i + N_i^S\langle\Delta lnP_{i_t}^S\rangle_i. \qquad (2.10.2)$$

The right side of this equation is an identity since it is the sum of all the tick price-changes after the prior day's closing price to the current day's closing price. One can sum the above equation over all days in the data set to obtain the price-change of IVE from the starting day to the ending day of the data set

$$\sum_{i=1}^{1245} ln(P_i^c/P_{i-1}^c) = \sum_{i=1}^{1245}\left\{N_i^B\langle\Delta lnP_{i_t}^B\rangle_i + N_i^S\langle\Delta lnP_{i_t}^S\rangle_i\right\} = ln(P_{1245}^c/P_0^c). \qquad (2.10.3)$$

In this equation 1245 is the total number of days in the data set and P_0^c is the closing price of the index on the day just prior to the start of the data set. This equation provides a useful check on the accuracy of the computed daily buyer (seller) price-changes per tick that the current results satisfied.

We now modify the terms on the right of equation (2.10.2) to introduce the sentiment variable in price-change units. As noted earlier the average magnitudes over the entire data set of the daily price changes per tick are essentially equal for buyers and sellers. Now we define the constant $A \equiv \left(\left|\langle\langle\Delta lnP_{i_t}^B\rangle_i\rangle\right| + \left|\langle\langle\Delta lnP_{i_t}^S\rangle_i\rangle\right|\right)/2 = 3.21\text{x}10^{-5}$, which is just the average of buyer and seller

[13] See here. The same Generalized Gamma distribution represents both the buyer and seller ones. The parameters of the latter reference were: $a = 1$, $d = 8$, and $p = 0.29$. The relation $f_{GG}\left(Y^{B(S)}/\sigma^{\Delta lnP_{i_t}^{B(S)}}\right) = \sigma^{\Delta lnP_{i_t}^{B(S)}} f_{GG}\left(Y^{B(S)}\right)$ was also used.

magnitudes. Multiplying the sentiment variable by this constant, one obtains the daily variable $A\left(N_i^B - N_i^S\right)$, which measures the daily bull-bear sentiment in units of daily price-change. This allows a direct comparison of daily bull-bear sentiment distribution to that of price-change which will be presented in Section 4.

3. Long-Term Intraday Sign-Correlation for Order-Flow and Price-Mobility

This section presents the long term sign-correlations for order-flow in Subsection 3.1 and those of price-mobility in Subsection 3.2.

3.1 Long-Term Intraday Sign-Correlations for Order-Flow

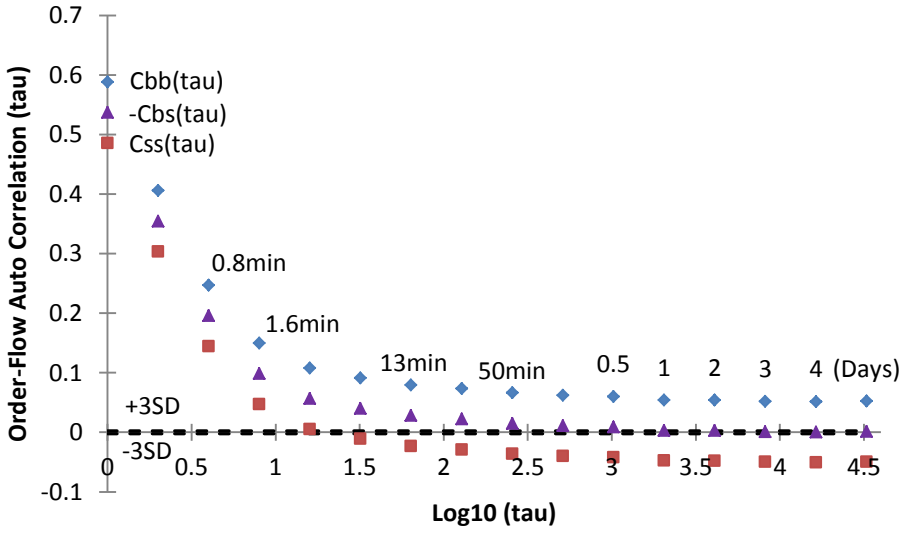


Figure 3.1.1 Auto correlations $C_{BB}^{LR}(\tau)$, $C_{SS}^{LR}(\tau)$, and the average sign correlation $\bar{C}_{OF}^{LR}(\tau) = -C_{BS}^{LR}(\tau) = \frac{1}{2}[C_{BB}^{LR}(\tau) + C_{SS}^{LR}(\tau)]$.

Figure 3.1.1 gives the auto correlations $C_{BB}^{LR}(\tau)$, $C_{SS}^{LR}(\tau)$, and the average order-flow sign correlation $\bar{C}_{OF}^{LR}(\tau) = -C_{BS}^{LR}(\tau) = \frac{1}{2}[C_{BB}^{LR}(\tau) + C_{SS}^{LR}(\tau)]$. These correlations are defined in Subsection 2.7. Labels appear on some correlations in the figure designating nominal[14] lagged

[14] The clocks times correspond to averages and are nominal in the sense that their corresponding standard deviations may be comparable to the averages or even larger.

clock times starting at about a minute and moving progressively higher out to several days. The three standard deviation (± 0.0024) dotted lines parallel the horizontal axis.

Let's consider the results obtained for order-flow sign correlations with the IVE data set in detail. The three sign correlations shown in the figure have significantly different asymptotic behaviors with increasing lags. The average order-flow sign correlations[15] $\bar{C}_{OF}^{LR}(\tau) = -C_{BS}^{LR}(\tau)$ decrease steadily with increasing lag time. At roughly the daily scale $\tau \sim 10^{3.3} \sim 2000$, this average correlation is 0.0034 and it approaches the three standard deviation level of 0.0023. The individual correlations of buy-buy $C_{BB}^{LR}(\tau)$ and sell-sell $C_{SS}^{LR}(\tau)$, on the other hand, initially decrease but then asymptotically approach constant values of ~ 0.5 and ~ -0.05, respectively. Note that if one correlation becomes positive it must necessarily be at the expense of the other which becomes negative. Hence the difference $C_{BB}^{LR}(\tau) - C_{SS}^{LR}(\tau)$ is twice the magnitude of the deviation of each one from the average of these correlations. It's remarkable that this difference is constant at all lags τ. In the current instance the buyers' correlations dominates these of sellers. Finally, as the average of the correlations approaches zero, the two correlations become equal in magnitude and have opposite signs.

Now consider the implications for the sentiment skew excess on the daily scale of the asymmetric behavior of the buyer and seller probabilities. Using the equations in Subsection 2.4, one btains the true probabilities of daily occurrence of buyer and seller trades and their standard deviations $P_B = 0.522(1 \pm 0.010)$ and $P_S = 0.478(1 \pm 0.010)$. Now the asymmetric Lee-Ready term with its sample-size standard deviation (see Table 4.1 in Subsection 2.4) is $e_{asym}^{LR} = -0.020 \pm 0.018 = -0.02(1 \pm .9)$. The following daily excesses are obtained for positive and negative standard deviations in the asymmetric Lee-Ready error:

$$\varepsilon_{daily}^{OF} = P_B - P_S$$

$0.046(1 \pm 0.19)$	e_{asym}^{LR} has positive standard deviation
$0.084(1 \pm 0.10)$	e_{asym}^{LR} has negative standard deviation
$0.065(1 \pm 0.14)$	Average

On the daily scale the IVE data set with its duration of about five years has $N_{tot} = 1245$ days and the excess $\varepsilon_{All,\,N_{tot}}^{BS} = 0.071(1 \pm 0.22)$. This daily-scale skew excess agrees with all values of the order-flow excesses in the summary listing above.

A similar more active buying than selling by high-frequency traders in response to company news announcements was found by Brogaard, Hendershott, and Riordan (2014). Their Figures 5 and 4 show high frequency traders' demand for liquidity (or flow of market orders for immediate

[15] This is the extremely persistent order-flow correlation first pointed out independently by Bouchaud, Gefen, Potters, Wyart (2004), and Lillo and Farmer (2004). Bouchaud, Farmer, and Lillo (2009) review such long memory processes.

execution) that they denote by HFT^D. Their Figure 5 shows a cumulative increase of buying demand in response to positive news that plateaus at $HFT^D \sim 400$. Their Figure 4 shows the corresponding cumulative selling demand with the plateau at magnitude $HFT^D \sim 350$. Hence their excess given by $\frac{400-350}{400+350} \sim 0.07(1 \pm 0.3)$[16] is comparable to the excesses above.

3.2 Long-Term Intraday Sign-Correlations for Price-Mobility

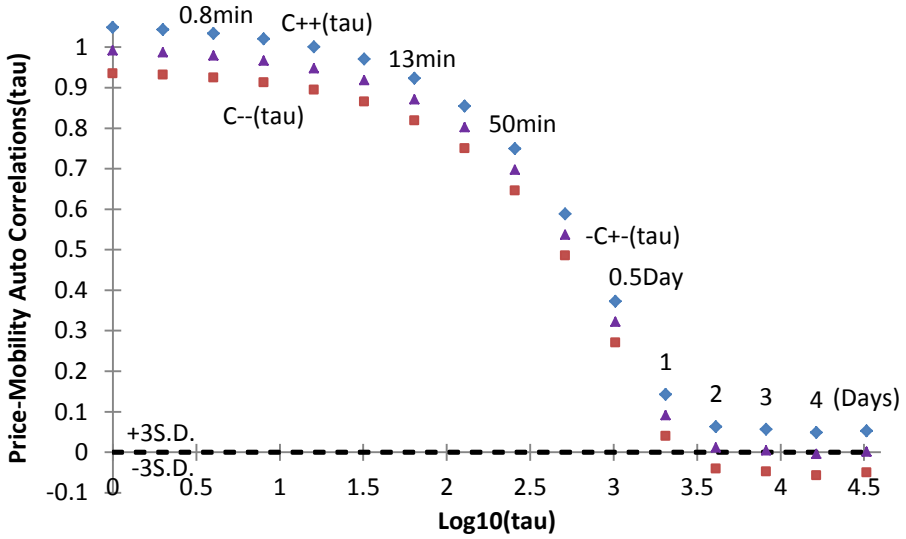


Figure 3.2.1 Individual price-mobility sign-correlations $C_{++}(\tau)$, $C_{--}(\tau)$, and the average price-mobility sign-correlation $\bar{C}_{PM}(\tau) = -C_{+-}(\tau) = \frac{1}{2}[C_{++}^{LR}(\tau) + C_{--}^{LR}(\tau)]$.

Figure 3.2.1 gives the individual price-mobility sign-correlations $C_{++}(\tau)$, $C_{--}(\tau)$, and the average price-mobility sign-correlation $\bar{C}_{PM}(\tau) = -C_{+-}(\tau) = \frac{1}{2}[C_{++}^{LR}(\tau) + C_{--}^{LR}(\tau)]$. These correlations are defined in Subsection 2.7. As in the prior figure, labels on some correlations designate nominal lagged clock times. The three standard deviation (± 0.0023) dotted lines parallel the horizontal axis. Surprisingly at first glance, the lowest five lag-times have $C_{++}(\tau) > 1$. However the correlations do sum to unity: $\frac{1}{4}[1 + C_{++}(\tau)] + \frac{1}{4}[1 + C_{--}(\tau)] + \frac{1}{2}[1 + C_{BS}(\tau)] = 1$, and the average sign-correlations behave as expected $\bar{C}_{PM}(\tau) < 1$.

As for the prior case for order-flow, the three price-mobility sign-correlations shown in the figure have significantly different asymptotic behaviors with increasing lag-times. The average

[16] The uncertainty here represents my estimates of accuracy in my determination of the HFT^D magnitudes with a ruler from their figures.

price-mobility sign correlations $\bar{C}_{PM}(\tau) = -C_{+-}(\tau)$ decrease steadily with increasing lags. Here at roughly the daily scale, this average correlation is 0.0091 and is significantly above the three standard deviations level of 0.0023, but it drops below the latter at the two day scale. Again as for the case of order-flow sign correlation, the individual price-mobility correlations $C_{++}(\tau)$ and $C_{--}(\tau)$ initially decrease but then asymptotically approach constant values of ~ 0.5 and ~ -0.05, respectively. The difference $C_{++}(\tau) - C_{--}(\tau)$ is again remarkably constant over all lag-times τ.

Let's compare the behavior of $\bar{C}_{PM}(\tau)$ to that of $\bar{C}_{OF}(\tau)$. The latter is the extremely persistent order-flow sign-correlation discussed in the literature and reviewed by Bouchaud, Farmer, and Lillo (2009). Figure 3 below shows the price-mobility and order-flow correlations for all lag-times where these correlations are greater than the three-standard deviation noise level. One sees that the price-mobility sign-correlations are much stronger than those of order-flow over essentially the entire range. At the daily lag the price-mobility correlations are a factor of 1.8 greater than those of order-flow. For lower lags at they are stronger by greater factors up to around 47. At lags of one and two days, respectively, the price-mobility correlations are 0.0112 and 0.005 and the order-flow ones are 0.0034 and 0.0033. In summary, both correlations have the same degree of persistence, i.e., they each remain statistically significant out to one or two lag-days. However the price-mobility correlation is much stronger over the whole range.

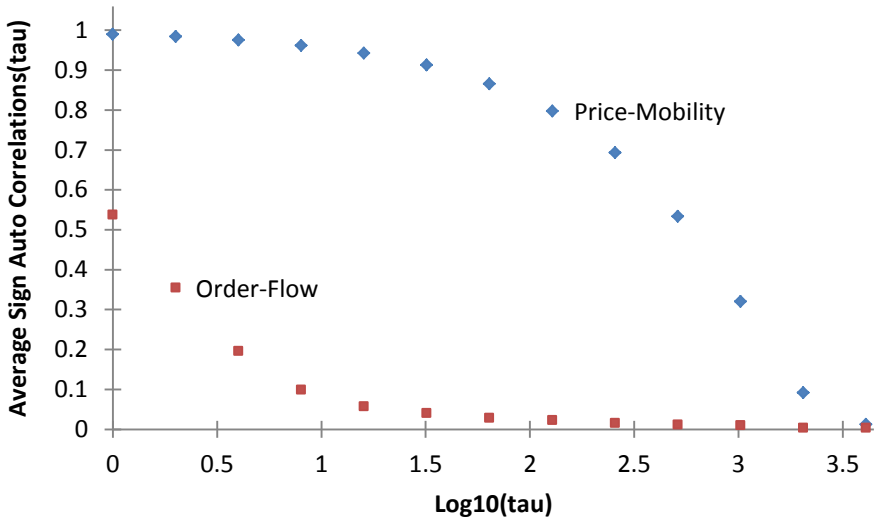


Figure 3.2.2 Price-mobility and order-flow correlations $\bar{C}_{PM}(\tau)$ and $\bar{C}_{OF}(\tau)$.

Using the relations in Subsection 2.4 one obtains the price rise and fall probabilities $P_+ = 0.523(1 \pm 0.004)$ and $P_- = 0.477(1 \pm 0.004)$, respectively. The price-mobility skew excess is thus given by $\varepsilon^{PM}_{daily} = P_+ - P_- = 0.046(1 \pm 0.17)$.

On the daily scale the IVE data set with its duration of about five years has $N_{tot} = 1245$ days and the excess $\varepsilon^{\Delta lnP^c}_{All,\, N_{tot}} = 0.053(1 \pm 0.27)$, which agrees with the daily price-mobility excesses.

The results on high-frequency trading presented by Brogaard, Hendershott, and Riordan (2014) are also relevant here. Their Figures 4 and 5 also show returns or price-changes due to buyer and seller activities in response to negative and positive news announcements, respectively. The response to positive news plateaus to $\sim 0.10\%$, while the magnitude of the response to negative news plateaus to $\sim 0.09\%$. Their excess is $\frac{0.10\% - 0.09\%}{0.10\% + 0.09\%} \sim 0.05\ (1 \pm 0.3)$ which again is comparable to our price-mobility excess. The uncertainty in their excess here represents my estimate of accuracy in measuring the "returns" magnitudes with a ruler from their figures.

4. Behaviors of Price-Change and Bull-Bear Sentiment Distributions with Increasing Scales

This section considers the behaviors of the price-change and sentiment distributions with increasing scales. At a given scale each distribution is uniquely specified by its summary parameters: the mean or drift $\langle N_j^X \rangle_s$, the volatility σ_s^X, the all-magnitude skew excess $\varepsilon_s^X(0)$, and the shape parameter defined by the ratio of volatility to mean absolute deviation $\sigma_s^X / \tilde{\sigma}_s^X$, where X denotes either price-change ΔlnP^c or sentiment $A(N^B - N^S)$. This section gives sentiment in units of price-change[17] as described in Subsection 2.10. Each summary parameter's scaling behavior with increasing scales is examined for both price-change and sentiment. We seek similar behaviors in each price-change and sentiment summary parameter with increasing scales that would indicate reciprocal effects between the two distributions.

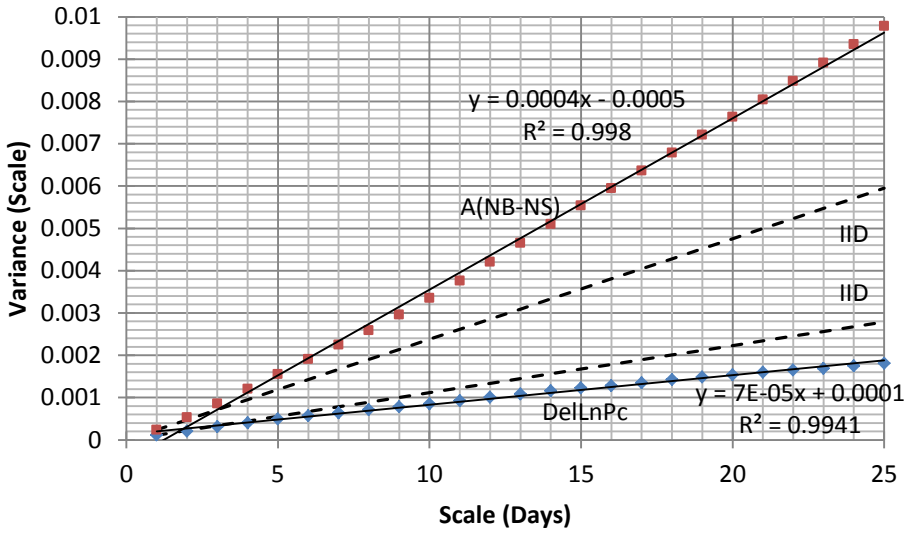
4.1 Scaling of Variances and Means

Figure 5.1.1 below shows the variance behaviors with increasing scales for each distribution. The variance at scale s of identical independent distributions (IID) is $(\sigma_s^X)^2 = s(\sigma_1^X)^2$, where X represents either price-change or sentiment. The latter are shown by dashed lines in the figure. The figure also shows each respective linear trend line and the R^2 value.

[17] The drift-free series elements in units of price-change here are $Z_{s,j}^{A\left(N^B - N^S\right)} = AZ_{s,j}^{\left(N^B - N^S\right)}$, where $Z_{s,j}^{\left(N^B - N^S\right)}$ represent the original sentiment elements at scale s, and Subsection 2.10 defines the factor A.

The variances increase linearly[18] with increasing scales. However, their linear increases deviate somewhat from IID behavior because of the presence of non-zero but negligible auto correlations for lag times at and above one day. These negligible but none-zero auto correlations are in the numerical variances at the higher scales causing the deviance from strict IID behavior. Appendix 1 derives expressions for variances in terms of auto correlations at various lag times on the daily scale. The appendix shows that adding the expressions in terms of the daily auto correlations gives results that equal the actual numerical variances.

Deviations from exactly IID behaviors are also present in the behaviors of means with increasing scales but they are significantly smaller than those of the variances.



Figures 4.1.1 Behaviors of price-change and sentiment variances: $\left[\sigma_s^{\sigma^{\Delta lnP^c}}\right]^2$ and $\left[\sigma_s^{A\left(N^B-N^S\right)}\right]^2$, respectively, with increasing scales for the IVE data, along with the corresponding IID variances (see text) denoted as dash lines.

[18] The IID like linear behavior of the variance for the logarithmic price-change variable is well known (cf. Figure 3 in Gopikrishnan, Plerou, Amaral, Meyer, and Stanley, 1999). The slightly steeper slope here is within their one standard deviation upper limit.

4.2 Scaling of Skew Excesses

Figure 4.2.1 below shows the behaviors with increasing scales of the skew excesses of the price-change and sentiment distributions, $\varepsilon_s^{\Delta lnP^c}(0)$ and $\varepsilon_s^{A(N^B-N^S)}(0)$, respectively. The central limit theorem mandates that IID distributions at increasingly higher scales evolve toward the normal distribution, which has a skew excess of zero. Hence the skew excesses of IIDs must decrease with increasing scales. These are shown by the dashed lines in the figure, with both decaying to zero by about the seven day scale.

The empirical skew excesses of price-change and sentiment distributions both generally increase or remain about level with increasing scales below about 16 days. The growth of the sentiment skew becomes somewhat anemic above the four day scale but it tends to trend higher from the daily scale through the 14 day scales. Such increases in skew excesses are contrary-to-IID behaviors. It's significant that both skews deviate from IID behaviors in the same direction in this region and they behave similarly at these lower scales. This suggests a close coupling between bull-bear sentiment and price-changes.

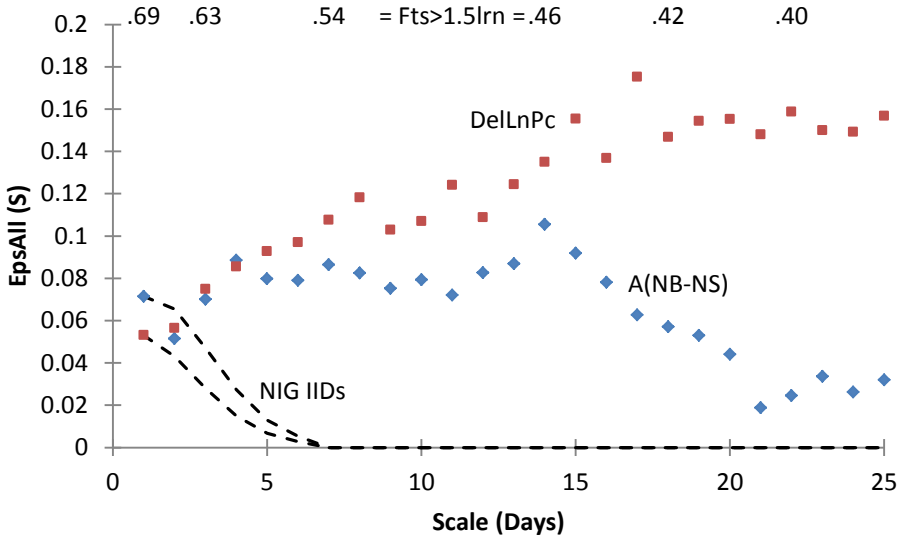


Figure 4.2.1 Behaviors with increasing scales of the all-magnitude price-change and sentiment skew excesses, respectively, $\varepsilon_s^{\Delta lnP^c}(0)$ and $\varepsilon_s^{A(N^B-N^S)}(0)$. Corresponding Normal Inverse Gaussian distribution's IID behaviors are shown in dashed lines. A measure of the fidelity $F_{TS>1.5LRN}(s)$ of the true signal above noise in the Lee and Ready data at several scales is given across the top of the figure (see text).

Above about the 16 day scale the skew excess for the bull-bear sentiment decays in a manner that parallels the NIG IID behavior, and reaches its minimum value in seven days at scale 21 lag

days. On the other hand, the price-change skew excess maintains an upwardly trending excess out to the seventeen day scale and then plateaus to a roughly constant value.

Appendix 2 gives a rough estimate of the error in sentiment due to the noise from the asymmetric Lee-Ready term. This appendix introduces a true signal fidelity parameter which represents the fraction of the series' components at each scale in which the true signal (TS) strongly dominates the Lee-Ready noise (LRN) $F_{TS>1.5LRN}(s)$. This signal fidelity parameter is given at several scales across the top of the above figure. Let's now re-examine the scaling behavior of the sentiment skew in light of this parameter.

One expects the bull-bear sentiment skew excess to increase at least as rapidly the price-change skew with increasing scales since the order-flow daily skew is even bigger than that of price-mobility. The sentiment excess in Figure 4.2.1 shows a systematic increase from the daily to fourteen-day scales similar to the rise of the skew excess of price-change, although the increase is not as robust as the latter one's. The weaker rise could be explained by the fact that only $\sim 70\%$ of its series components dominate the noise strongly as shown by the fidelity parameter $F_{TS>1.5LRN}(1)$. This parameter has values that vary from ~ 0.7 at the daily scale to about ~ 0.5 at the fourteen day scale. Thus at the daily scale only $\sim 70\%$ of the data set has strong signals relative to the noise, and this fraction decreases to $\sim 50\%$ at the fourteen day scale. Beyond the fourteen-day scale the excess shows a steady and systematic decline downward, while the true signal fidelity drops down to ~ 0.40 at the twenty-two day scale. Thus it seems that as long as the signal fidelity is above ~ 0.5 the excess tends to rise as expected, but once it drops below the latter value the noise overwhelms the true signal and the excess tumbles downward with increasing scales. Hence as long as more than about half of the data set's constituents have a very strong true signal the approach using the Lee-Ready classification method for classifying buyers and seller appears to give roughly valid distributions' skews.

4.3 Scaling of Shape Parameters

Figure 4.3.1 below shows the behaviors with increasing scales of the price-change and sentiment shape parameters: the ratios of volatility to absolute deviation. The central limit theorem mandates that IID distributions evolve toward the normal distribution, hence IID shape parameters decrease to $\sqrt{\pi/2}$. Dashed lines in the figure show IID behavior of the respective normal inverse Gaussian distributions. The shape parameter for the NIG IIDs with initial summery parameters corresponding to the empirical daily values of the price-change and sentiment variables decay to the Gaussian limit at a scale of around 6 days.

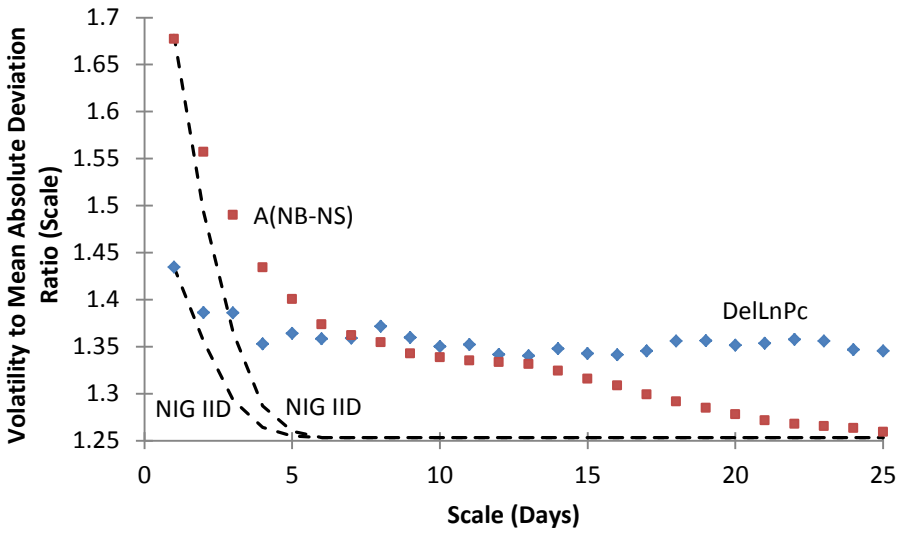


Figure 4.3.1 Behaviors of price-change and sentiment scaling parameters: $\sigma_s^{\Delta lnP^c}/\tilde{\sigma}_s^{\Delta lnP^c}$ and $\sigma_s^{A(N^B-N^S)}/\tilde{\sigma}_s^{A(N^B-N^S)}$, respectively. Dashed lines show corresponding NIG IID behaviors.

The empirical decreases of both $\sigma_s^{\Delta lnP^c}/\tilde{\sigma}_s^{\Delta lnP^c}$ and $\sigma_s^{A(N^B-N^S)}/\tilde{\sigma}_s^{A(N^B-N^S)}$ in the figure are markedly different from those of NIG IIDs. The sentiment shape parameter starts from a value above that of the price-change but decays more quickly and reaches the latter's value by about the seven day scale. So both empirical shape parameters first decrease somewhat like IIDs but then stop decreasing and the rate of decrease towards Gaussian stalls. Then they exhibit similar behaviors above the seven-day scale where they are plateauing at roughly the same levels. These levels are significantly above the Gaussian limit out to about the 16-day scale. The scale range of initial decrease and plateauing is roughly three times the range of the NIG IID decays. However, above about the 15 day scale the sentiment shape parameter begins to decay steadily. At the scale of ~25 days, the sentiment scale parameter decays to the Gaussian value, while the price-change shape parameter continues to plateauing. The latter decay again is due to the Lee-Ready noise that was discussed in Subsection 4.2 on skew excesses.

5. Auto Correlations of Daily Buyer and Seller Totals

This section examines the behavior of auto correlations of buyer and seller daily totals with increasing lag times. It first discusses the general decay of the correlations with increasing lags summarizing relevant results from Appendices 2 and 3. It infers that the general decay characteristics are probably due to period of the IVE data set that contains the great recession. It

then focuses on the two tiered decay of buyer over seller decays and relates this feature to the persistent order-flow correlations.

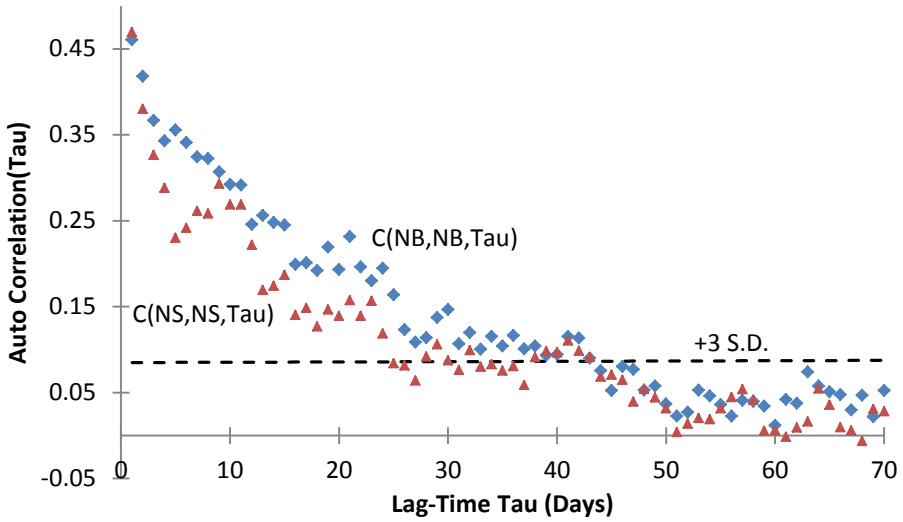


Figure 5.1 Auto correlations of buyer and seller daily totals $C(N^B, N^B, \tau)$ and $C(N^S, N^S, \tau)$, respectively.

The figure above gives the true auto correlations for the buyer and seller daily totals. The correlations exhibit a rapid decay rate that brings them down to noise level[19] in about thirty to forty lag days. Appendix 3 considers the auto correlations of the volume-weighed buyer and seller daily totals. It shows the latter correlations have decay rates that are essentially parallel to the correlation of the total daily volume. This indicates that the volume-weighed correlations are probably in part due to the Lee-Ready noise[20] which is proportional to total daily volumes. Appendix 3 uses the total volume correlations of the S&P 500 index to argue that the decay rates of the volume-weighed buyer and seller auto correlations derive their characteristics from the specific period of our data set. This period contains the great recession where the S&P 500 Index's daily volume correlations decay much more rapidly than they do for more typical market periods. The decays of the buyer and seller daily trade totals in the above figure likely also contain significant contributions from the Lee-Ready noise. The rapid decays of correlations in the figure are thus probably the signature of the great recession, while more typical market periods would have significantly slower ones.

A striking feature of the above figure is that the buyers auto correlation track lies systematically above the one of the sellers. This difference in rates of correlation decay is probably due to the difference in probabilities of buyer and seller trade occurrences due to the persistent order-flow correlations. The magnitude difference in strength of these correlations is proportional to the

[19] The noise level is $3/\sqrt{N_{tot} - \tau}$, where $N_{tot} = 1245$, the total days of data, and τ is the lag time.
[20] See discussions on scaling of the all-magnitude skew in Subsection 4.2 and of the Lee-Ready noise in Appendix 2.

difference in the true probability products $P_B^2 - P_S^2 = P_B - P_S \cong 0.065$ where the latter value is the daily order-flow skew excess determined in Section 3. Averaging the empirical differences plotted in the above figure using the first twenty-five lag days that are mostly above the noise level gives the empirical value of $\sim 0.05 \pm 0.03$, which agrees with the daily order-flow excess.

6. Summary and Discussion

Section 3 presented the intraday sign-correlations of order-flow and introduced those of price-mobility. The respective average sign-correlations[21] decay to noise above the nominal one day lag, with the price-mobility correlations being much stronger on all smaller lags. The individual sign-correlations remain way above the three standard deviation level out to several nominal days of lags. A remarkable feature is that the difference between buyer and seller correlations for order-flow and the corresponding difference for price-mobility are independent of the lag times. At the longest lags the individual correlations of both order-flow and price-mobility are equal in magnitude but have different signs. Section 3 also determined the probabilities of occurrence on the daily scale of a buyer or a seller trade. It also determined the daily probabilities of occurrence of a daily price rise and a price fall. The difference between the latter probabilities was used in the introduction to show how these correlations create the positive skew in the price-change distribution.

Section 4 examined the behaviors of the price-change and sentiment distributions' summary parameters. It found that the distributions' means and variances behave like IIDs. The price-change and sentiment skew excesses show definite behaviors that are contrary to those of NIG IIDs, with both deviating in the same direction. Both empirical skews initially increase. While the price-change skew continues its increase, the sentiment skew plateaus and then collapses. The latter sentiment skew behaviors are due to the growth of noise in the sentiment with increasing skew due to the asymmetric Lee-Ready errors in the buyer (seller) classification procedure. The parallel deviations of the skews parameters from IID behavior for price-change and sentiment are due to extremely persistent intraday sign-correlations of order-flow and price-mobility. The price-change and sentiment shape parameters' behaviors also deviate strongly from those of NIG IIDs. The deviation here is in the rate of approach to the limiting normal distribution's value. Both initially fall but at different rates, and their values approach each other. Then they both plateau out to about the sixteen-day scale. The similar behaviors of the two are probably due to the intraday correlation. Such similar behaviors are also expected since price-change affects sentiment and visa-versa. Subsequently the sentiment shape parameter

[21] The extremely persistent sign-correlations in the literature are the average correlation of order-flow in our formulation.

collapses to the Gaussian limit while that of price-change continues to plateau. The collapse of the sentiment shape parameter again is due to the Lee-Ready errors' growth at the higher scales.

Section 5 presented the buyer and seller daily totals auto correlation. These correlations decay to noise in about thirty to forty lag days. It summarized the arguments of Appendix 3 that the decay is likely a characteristic of the IVE data set's period which contained the great recession, and may be due to significant noise from the Lee-Ready errors. The buyer and seller daily auto correlations also exhibit tiered decay trajectories with buyers lying above sellers trajectories. The intraday order-flow sign-correlations can explain this tiered decay.

The intraday sig-correlations of price-mobility create the positive skew in the daily price change distribution as was shown in the introduction. This positive daily skew then gives rise to the growth of skews with increasing scales as follows. To form the two-day scale variable one adds two sequential daily variables. Since each of the individual days added together on average contain the daily bias to positive skew, the process of addition amplifies at the two-day scale the initial positive daily bias, giving a stronger bias to positive skew at the two-day scale. Such amplification continues to act at each transition from scale s to scale $s + 1$. Hence the positive bias in the daily price-changes assures that the skew tends to continue its increase at successively higher scales. The same considerations apply to order-flow correlations' introduction of positive bias in the daily sentiment. It creates a positive daily sentiment skew excess that tends to continue to increase at successively higher scales.

Let's now address the source of these extremely persistent correlations. Toth, Palit, Lillo, and Farmer (2015) show that order-splitting drives the long-term memory exhibited by order-flow for trading times up to a few hours. These authors note that their results are limited to times below a few hours because at longer ones they "begin to lose statistical significance in [their] results". Their study uses the average of the buyer and seller sign correlations $\bar{C}_{OF}$ rather than the individual ones C_{BB} and C_{SS}[22]. The average loses statistical significance much more rapidly than the latter two (see Figure 3.1.1 in Subsection 3.1). The use of the latter two correlations possibly could extend their conclusion that order-splitting causes the long-term memory in order-flow from a few hours to several days.

The source of the price-mobility sine-correlations is also order-splitting. The existence of these correlations can be understood as follows. The price-mobility variable is the sign of the difference between an intraday trade price and the prior day's closing price. The intent of order-splitting is to limit the price impact of large orders and suppress big price moves. To the extent it's successful, order-splitting anchors the intra-day price-mobility to the prior day's closing price, creating the observed intraday long-term sign correlations.

[22] These correspond, respectively, to this study's $\bar{C}_{OF}^{LR}$, C_{SS}^{LR}, and C_{BB}^{LR}, but they are not limited by the Lee-Ready classification method.

Closing this discussion, let's conjecture on the sources behind the negative skew excess at high magnitudes for price-change as shown in Figure 1 of the introduction. The figure shows that above $\sim 1.4\sigma_S^{\Delta lnP^c}$ the skew excess is decidedly negative at both the daily and twenty-five day scales. The skew excesses at high magnitudes of sentiment distribution are also negative at high magnitudes for the lower scales of the distributions discussed in Section 4. Thus the following conjectures also apply to the high magnitude skews of sentiment.

The negative skew excesses at high magnitudes are suggestively in line with prospect theory of behavioral economics (cf. Kahneman 2011). Prospect theory holds that human behavior magnifies the fear of potential losses to a much greater degree than it does the attraction to potential gains. Thus it implies that agents initiating sell orders should act more quickly than those initiating buy ones, and there should be more active sellers than buyers when large magnitude price swings ensue. The quicker selling by a larger pool of sellers should give greater daily trade totals for sellers than those of buyers at high magnitudes. The fact that the stakes are bigger for high magnitude price-changes should enhance the trader's natural human fear of losses and cause them to sell more quickly in larger order lots that have more price impact. Thus prospect theory helps to explain the negative empirical skews of sentiment at high magnitudes. Additionally, regulations and business practice tend to encourage more rapid selling for large and sustained price moves since fund managers must issue quarterly reports on holdings. As noted by Keim and Madhavan (1995), once the decision to sell is made, the institutional trader is penalized more by selling a sinking stock too slowly than by buying a surging one too slowly. This is because the slow selling incurs a measurable accounting loss while the slow buying produces an unobservable opportunity loss. The above reasoning thus implies that the skews should be negative at the high magnitudes for both price-change and bull-bear sentiment.

The author thanks Damon Larson for editing parts of this manuscript.

Appendix 1: Higher Than Daily Scale Variances and Daily Autocorrelations

This appendix first summarizes the derivation of the relation between higher than daily scale variances and daily auto correlation. It then applies the results to show how daily correlations contribute to higher scale variances.

Denote a series' elements at scale s by ${Y^X}_{j,s}$ where X represents any one of the following variables N^B, N^S, $A(N^B - N^S)$, or ΔlnP^c and the total of series elements by N_{tot}. One can readily show that the mean or drift at each scale is given by $\langle {Y^X}_{j,s} \rangle \cong s\langle {Y^X}_{j,1} \rangle$. The approximate equality approaches full equality as $s/N_{tot} \to 0$.

Using $\left(\sigma_s^X\right)^2 = \left(\sum_{s=1}^{s} Z_{j,s}^X\right)^2$, where $Z_{j,s}^X = {Y^X}_{j,s} - \langle {Y^X}_{j,s} \rangle$ are the drift-free sequence elements, one can show that the variance $(\sigma_s^X)^2$ satisfies the following recursive relations for $s > 1$

$$\frac{\left(\sigma_s^X\right)^2}{\left(\sigma_1^X\right)^2} \cong \frac{\left(\sigma_{s-1}^X\right)^2}{\left(\sigma_1^X\right)^2} + 1 + 2\sum_{\tau=2}^{s} C_{s=1}(Y^X, Y^X, \tau - 1), \qquad \text{(A1.1)}$$

where $C_{s=1}(Y^X, Y^X, \tau - 1)$ is the auto correlation at lag time $\tau - 1$ on the daily scale. Again the approximate equality approaches full equality as $s/N_{tot} \to 0$.

Now using this recursive relation and the step function $\Theta(y)$, where $\Theta(y) = 1$ for $y \geq 0$ and 0 otherwise, one derives the relation for higher scale ($s > 1$) variances and the daily auto correlations

$$\frac{\left(\sigma_s^X\right)^2}{\left(\sigma_1^X\right)^2} \cong 1 + \sum_{\tau=2}^{s}\left[\sum_{j=2}^{s} 2\Theta(j - \tau)\right] C_{s=1}(Y^X, Y^X, \tau - 1). \qquad \text{(A1.2)}$$

Now consider the results obtained from the last equation for the variables: N^B, $A(N^B - N^S)$, and ΔlnP^c (the N^S results are essentially the same as those of N^B) for the IVE data set. Let's compute the variances from this equation using only the auto correlations at lag times between 1 and 11 days to examine effects of the neglected higher lag auto correlations. This "correlation variance" will be compared to the actual variance for scales from 1 to 25 days.

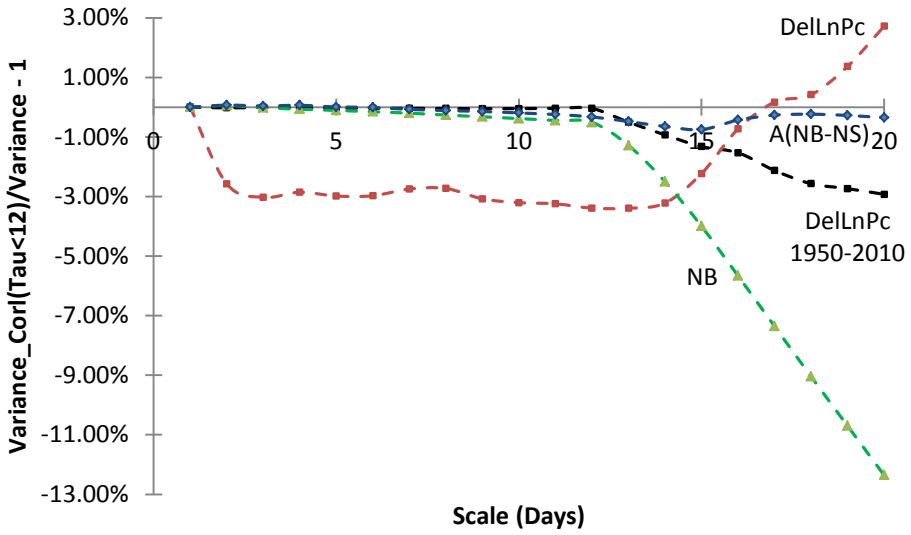


Figure A1.1 Percentage differences between the "correlation variances" computed using only auto correlations with lag times $\tau < 12$ and the actual variances for variables N^B, $A(N^B - N^S)$, and ΔlnP^c, where the latter is shown for both the IVE data set and the much longer one of Figure 1.1.

Figure A1.1 above shows the percentage differences between the "correlation variances" computed with only auto correlations at lag times $\tau < 12$ days included and the actual variance in colored symbols for the different variables for the IVE data set. It also shows this difference in black symbols for the variable ΔlnP^c for the much larger data set of Figure 1.1 in Section 1.

One sees that below lags of 12 days where all needed correlations are included the differences between the "correlation variance" and the actual one are small and roughly constant. Above lags of 12 days the differences can become significant due to the needed but not included correlations. Note the significand decrease in the discrepancy for price-change below 12 days lag in going from the IVE data set to the much larger data set of Figure 1.1. The effects of the needed but neglected correlations are most forcefully demonstrated for the case of N^B where strong correlations far above the levels of noise are present (cf. Figure 5.1 in Section 5). The actual variance increases rapidly above 12 day lags while the growth in "correlation variance" is stunted due to the neglected correlations. Thus the difference between the two increases dramatically with increasing lags.

Appendix 2: A Rough Estimate of Error in Bull-Bear Sentiment Due to the Asymmetric Lee-Ready Term

This appendix identifies the Lee-Ready noise and defines the true signal fidelity parameter which is a rough measure of the error due to this noise.

The initial anemic growth and subsequent collapse of the sentiment all-magnitude skew in Figure 4.2.1 of Subsection 4.2 is due to artificial mixing between buyer and seller trade totals introduced by the asymmetric Lee-Ready error term. The relation (see Subsection 2.4) between the true sentiment and the Lee-Ready difference and sum of buyer-seller totals at each scale is

$$N_{j,s}^{B} - N_{j,s}^{S} = \frac{1}{(1-2e_{sym}^{LR})}\left(N_{j,s}^{B,LR} - N_{j,s}^{S,LR}\right) + \frac{e_{asym}^{LR}}{(1-2e_{sym}^{LR})}\left(N_{j,s}^{B,LR} + N_{j,s}^{S,LR}\right). \qquad \text{(A2.1)}$$

In absence of the rightmost term the true sentiment would be proportional to the Lee-Ready sentiment, and the true distribution and the Lee-Ready one would be similar. The asymmetric error term however introduces a term proportional to the sum of the Lee-Ready buyer and seller totals that obviously mixes the two and creates noise. This noise can become sufficiently large to overwhelm the true signal contained in the first term on the right in the above equation.

To get a rough estimate of the true signal strength relative to noise consider the following. The strength of the true signal in the data at each scale is determined by how many of the series constituents have the first term on the right of the above equation significantly dominating the second one, i.e. we count the number of data points that satisfy

$$\frac{1}{(1-2e_{sym}^{LR})}\left|N_{j,s}^{B,LR}-N_{j,s}^{S,LR}\right|>1.5\frac{e_{asym}^{LR}}{(1-2e_{sym}^{LR})}\left(N_{j,s}^{B,LR}+N_{j,s}^{S,LR}\right). \quad \text{(A2.2)}$$

The factor of 1.5 in the above equation was chosen emphasize the most strongly dominating true signal constituents. Dividing this result by the total number of data points defines the signal fidelity parameter $F_{TS>1.5LRN}(s)$ which is the fraction of total data where the true signal (TS) exceeds the Lee-Ready noise (LRN) by a factor of 1.5. Table A1.1 gives $F_{TS>1.5LRN}(s)$ at various scales. The latter is used to understand the scaling behaviors of the skew and shape parameters in Section 4.

Table A1.1 Fractions of Data at scale s with true signal > 1.5 times the Lee-Ready noise.

$s=$	1	2	7	14	18	22
$F_{TS>1.5LRN}(s)=$	0.692	0.631	0.542	0.463	0.425	0.404

Appendix 3: Daily Volume-Weighed Correlations' Mimicry of S&P 500 Index's Total Volume Correlations in the Period of the Great Recession

This appendix first discusses the correlations of the volume-weighed daily buyer and seller trade totals and their differences from the correlations of trade totals in Section 5. It then focuses on the similarity of the former correlations to those of the total volume. This appendix subsequently compares the IVE total volume correlations to the S&P 500 Index's total volume correlations in different periods of comparable durations to the IVE data set. From this comparison it infers that the correlations of the buyer (seller) daily trade totals and volume-weighed totals probably reflect the period of the IVE data set that contains the great recession.

Let's now consider the correlations of the daily volume-weighed buyer and seller daily totals. Figure A3.1 below shows the auto correlations of the true volume weighed daily totals. The higher tracking trend by buyers over sellers seen in the figure with daily trade totals of Section 5 is obliterated for the volume auto correlations. The average of differences between the buyer and seller correlations for lag-times of one day to around fifteen days that include only correlations above noise levels is only ~ 0.02 here. This is about a factor of 20 below that of the average difference for daily totals in Section 5.

This disappearance is probably due to the following. For each individual trade the buyer or seller designation is a binary one, i.e. either a buy or sell. These individual binary buys or sell are then summed to determine the respective daily buyer and seller totals.

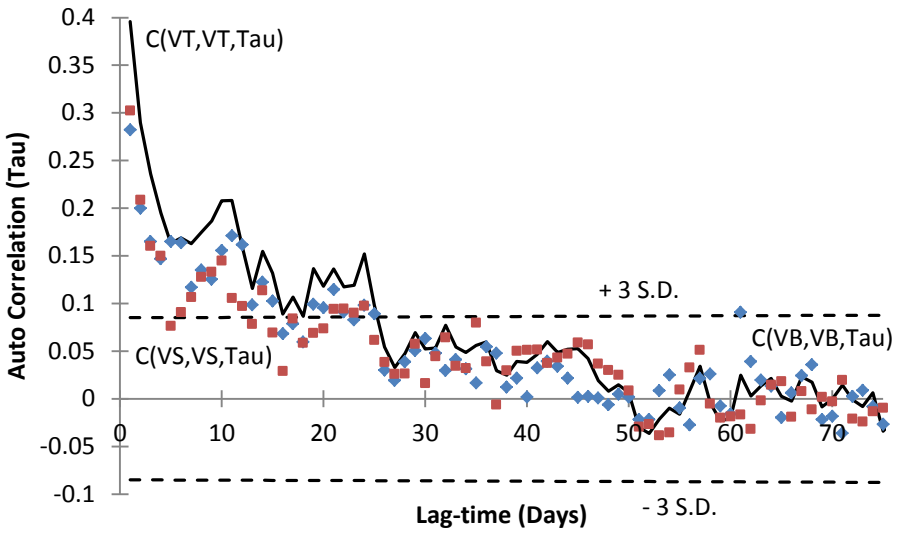


Figure A3.1 Auto correlations of buyer and seller daily volume totals $C(V^B, V^B, \tau)$ and $C(V^S, V^S, \tau)$, respectively. The solid line gives correlations of total volume $V^T = V^B + V^S$.

On the other hand each trade's volume is not a simple buy (sell) but varies in the number of shares traded. Thus the daily sums over buyer (seller) trades are over variable numbers of shares for each trade. This added degree of freedom in the share volume designation thus decreases the strength of the correlations, giving weaker correlation in the above figure. The decays of the correlations for daily volumes seem to parallel those of the trade totals but with strengths that are roughly two thirds that of the latter.

The above figure also shows the auto correlations of the total daily volume as a black solid line. Note that these total daily volume correlations have essentially the same behavior as the individual buyer and seller ones. This may indicates that the buyer and seller auto correlations are associated with the Lee-Ready noise[23] which is proportional to the total volume. Let's now compare the correlations decay of total volume in the IVE data set which begins on September 28, 2009 and ends on September 8, 2014 with the duration of about five years to the volume correlations of the S&P 500 Index in several comparable periods of duration.

Specifically let's use the data for of the S&P 500 Index's on daily prices and volumes in the period spanning 1/2/1991 to 12/31/2015 from Yahoo! Finance. Let's examine the following five-year periods in this data set: 2011 to 2015, 2006 to 2010, 2001 to 2005, 1996 to 2000 and 1991 to 1995. Note that the first period roughly corresponds to the period of our IVE data set in the above figure. The auto correlation of the total daily volumes for each of these five year periods were determined. Figure A3.2 below shows the results labeled by each period's end year. The correlations of our IVE data set are also shown in the figure.

[23] See discussion on the scaling of the all magnitude skew in Subsection 4.2 and the preceding appendix.

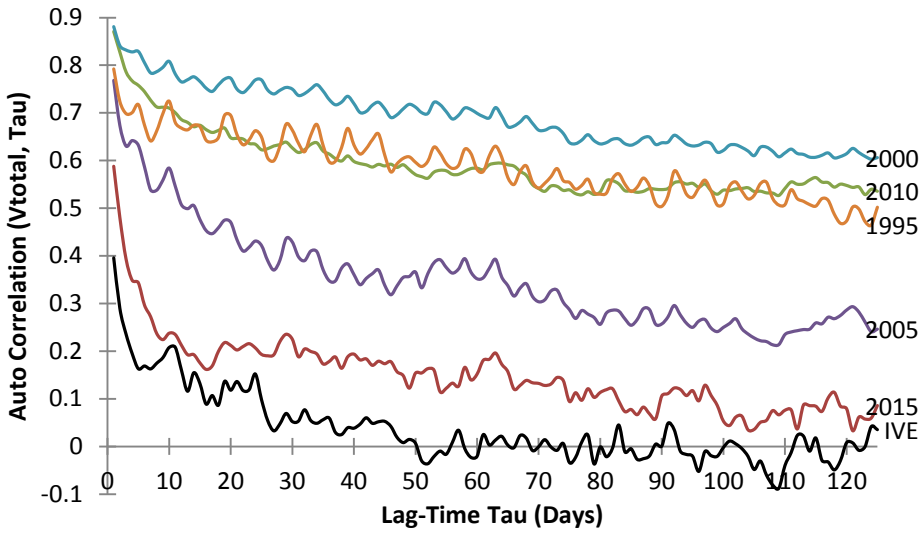


Figure A3.2 Auto correlations of daily volumes of the S&P 500 index in five different periods labeled by the end year. The IVE volume auto correlations are also shown.

One sees that peak correlations and subsequent decay rate can be very different from typical ones for some periods. The most typical correlation decay is rather slow as given by the periods ending in 1995, 2000, and 2010. The much more rapid decays are in the periods which end in 2005 and 2015. The 2005 period included the dot com crash, while the 2015 period had the great recession. This suggests that the dot com crash may have speeded up the decay rate of correlations in the 2005 period relative to typical ones. The great recession caused an even more rapid rate of correlation decay in the 2015 period. Note that for the 2015 and the IVE correlations where the periods roughly overlap, the rapid decays are roughly paralel with the track of the S&P 500 Index's correlations lying somewhat above that of the IVE ones. This suggests that the great recession has influenced both of these in a similar way.

One other consideration in support of the prior paragrph's inferrence of the great recession's influence in the latter five year period is the following. For the entire data set of the S&P 500 Index over the period from 1/3/1950 to 1/11/2016, the peak correlations at one lag day is 0.98 and the minimum of all lagged correlations considered is 0.93. Here the effects of the great recession and the dot com crash as well as other market falls are overwhelmed by the more typical calm market periods over the entire data set. Thus the effects of the great recession on volume correlations appear only in smaller data sets with durations around five years.

Finally one expects that the auto correlations of volume-weighed daily totals as well as the daily trade totals of buyer and seller show similar varying rates of decay for different periods as those of the S&P 500 Index's volume correlations in the figure above. This expectation is based on the similar behaviors of these correlations during the period of the IVE data set. Hence one infers that the rapid decays of the auto correlations of daily volume-weighed trade totals in Figure A3.1

above and of the daily totals in Figure 5.1 of Section 5 are probably the signiture of the great recession.

Referrences

Barndorff-Nielsen, O. E., 1998, “Processes of Normal Inverse Gaussian Type,” Finance Stochastics, Vol. 2, (1998) p. 41

Bouchaud, J-P., Farmer, J. D., Lillo, F., 2009, "How Markets Slowly Digest Changes in Supply and Demand," Handbook of Financial Markets: Dynamics and Evolution, p. 57

Bouchaud, J-P., Gefen, Y., Potters, M., and Wyart, M., 2004, “Fluctuations and response in financial markets: The subtle nature of random price changes", Quantitative Finance, Vol. 4(2), p. 176

Brogaard, J., Hendershott, T., and Riordan, R., 2014, “High-Frequency Trading and Price Discovery,” Review of Financial Studies, Vol. 27, p. 2267

Forsberg, L., 2002, “On the Normal Inverse Gaussian Distribution in Modeling Volatility in the Financial Markets,” Ph.D. Thesis, Uppsala Universitet

Gopikrishnan, P., Plerou, V., Amaral, L. A. N., Meyer, M., and Stanley, H. E., 1999, “Scaling of the Distribution of Fluctuations of Financial Market Indices,” Phys. Rev. E, Vol. 60, p. 5305

Kahneman, D., 2011, “Thinking, Fast and Slow” (Farrar, Straus and Giroux, New York 2011)

Keim, D. B., Madhavan, A, 1995, “Anatomy of the Trading Process: Empirical Evidence on the Behavior of Institutional Traders,” Journal of Financial Economics, Vol.37, p. 371

Lee, C. and Ready, M., 1991, “Inferring trade direction from intraday data,” Journal of Finance 46, p. 733

Lillo, F., and J. D. Farmer, 2004, "The long memory of the efficient market", Studies in Nonlinear Dynamics & Econometrics Vol. 8, Issue 3, p. 1

Odders-White, E., 2000, “On the Occurrence and Consequences of Inaccurate Trade Classification,” Journal of Financial Markets, Vol. 3, p. 259

Press, W. H.,Teukolsky, S. A., Vetterling, W. T., Flannery, B., P., 2002, “Numerical Recipes in C: The Art of Scientific Computing (Second Edition),” (Cambridge University Press, New York, 2002)

Toth, B., Palit, I., Lillo, F., and Farmer, J. D., 2015, “Why is equity order flow so persistent?” Journal of Economic Dynamics and Control, Vol. 51, p. 218