The Impact of High-Frequency Trading on Stock Market Liquidity Measures

Soohun Kim and Dermot Murphy*

This Draft: July 2011

Abstract

We examine potential misspecification in four commonly used empirical market microstructure models by comparing the model-implied effective spreads with the corresponding observed effective spreads for the S&P500 exchange traded fund (SPY) for the period 1997-2009. While these models, on average, underestimate the effective bid-ask spread by only 7 percent during 1997-2006, that figure rises sharply to 41 percent during 2007-2009. We conjecture that the increase in magnitude of underestimation in the latter period is caused by significant changes in patterns of trading after 2006: The average trade size declined from 2,700 shares during 1997-2006 to 400 shares during 2007-2009, and at the same time, the average number of consecutive buys or sells has increased from 4 to 12. This suggests that with the advent of high-frequency trading, it has become increasingly common to split up large orders into many smaller-sized orders and direct them to different trading venues. If we aggregate consecutive buys or sells into single trades by summing over volumes and size-weighting the prices, the underestimation of the effective spread is significantly reduced.

*Soohun Kim and Dermot Murphy are from the Kellogg School of Management at Northwestern University. We thank Ravi Jagannathan, Bob Korajczyk, Mike Fishman, Bob McDonald, Merrell Hora, Kevin Crotty, Andrei Jirnyi, Jingling Guan, Chang Joo Lee, and participants at the financial econometrics doctoral seminar at Kellogg School of Management for helpful comments and suggestions. All errors are our own. Corresponding author: dermot-murphy@kellogg.northwestern.edu.

1. Introduction

Market microstructure models are commonly used to provide estimates of the costs of executing a trade, when high frequency data on trades and quotes are available. The costs are useful in determining the liquidity of a stock over some time period. Numerous models based on transaction-by-transaction price changes are able to estimate the costs of trading by decomposing price changes into two components: the fixed cost component, usually defined by the bid-ask spread, and the variable cost component, which is typically increasing in the number of shares contained in the trade. These models typically assume a quote-driven market, where a single market-maker observes order flow and sets prices accordingly.

The variable cost component of a single trade is typically identified by the size of that transaction. An explanation typically offered by market microstructure models is that a large trade will command a higher price increase because market makers identify an adverse selection problem contained in such a large order (Kyle (1985)), or because market makers charge more for being forced to deviate from their optimal inventory position (Stoll (1978)). Therefore, while a small-sized order will simply incur a transaction cost equal to the bid-ask spread, a large-sized order will incur both the cost of the bid-ask spread and the additional cost charged by the market maker for having to deal in such a large share quantity.

Nowadays, order-driven markets are more prevalent, and all traders can act as a market maker because of their ability to place limit orders of specified prices and quantities on the limit order book. The limit order book still acts a "central market maker" in that a supply schedule is posted, and liquidity-demanding traders arrive to have their trades executed. A fixed transaction cost will still be incurred equal to the bid-ask spread, and additional costs will be incurred via large trades, due to the fact that the supply schedule is upward-sloping - this is akin to "walking up" the limit order book. One could argue, then, that empirical market microstructure models from the past, which are based on a quote-driven market, could still be applicable today because the limit order book could be substituted for the centralized market maker.[1]

However, a major problem exists in fitting past empirical market microstructure models to current data. In order to identify the variable costs of trading, the econometrician requires variation in the size of a trade. This was not a problem in the past - for

[1]However, Goettler et al. (2005) argue that measures of transactions costs may differ in a dynamic limit order book, as opposed to intermediate markets, due to price-setting by strategic traders.

example, in 1997, the average size of an individual trade for the S&P500 ETF (SPY) is 5,600 shares, with a standard deviation of 22,000 shares. That is, there is significant variation in the size of a single transaction so that the variable costs in trading a certain number of shares is identified. However, in 2009, the average size of a trade is 400 shares, with a standard deviation of only 5,100 shares. With significantly less variation in the size of a share transaction, it becomes more difficult to identify the variable cost in trading a certain number of shares. Without identification, the variable cost of a share transaction would be significantly underestimated.

Individual transactions occur in significantly smaller quantities nowadays, but the number of consecutive transactions has significantly increased. For example, in 1997, one would observe an average of two consecutive buy or sell transactions in the SPY, while in 2009, one would observe an average of twelve consecutive buy or sell transactions; in 2009, it was even possible to observe a sequence of over 100 consecutive buy or sell transactions. This corroborates the story that traders are dividing up their large orders into smaller orders, most likely because they want to avoid revealing the information content contained in their demand, or simply to minimize price impact.

The ability to trade in small quantities has increased over time for many reasons: the decrease of the minimum bid-ask spread from one-eighth of a dollar to one-sixteenth of a dollar in 1997, and another decrease to one penny in 2001[2]; increasingly sophisticated technology that allows investors, both individual and institutional, to easily trade in small quantities at a low cost; increased ease-of-access to the market, which increases the participation of investors with lower capital; and the advent of algorithmic trading, which allows for sophisticated strategies that perform many trades in smaller quantities over time, thus helping to hide the information content contained in the aggregate trade. [3] The latter is especially popular among brokerage firms that execute large trades on behalf of their clients with minimal price impact.[4]

While trading behavior has changed over time, both in the size of an individual order (which has become smaller) and the number of consecutive buy or sell orders

[2]Bessembinder(2003) investigates trade execution costs and market quality after 2001 change to decimal pricing; Stoll and Schenzler(2006) show the effect of decimalization on the quoted and effective spreads.

[3]Trading in smaller quantities also avoids drawing the attention of predatory traders, who trade to exploit the need of other investors to reduce their positions (see Brunnermeier and Pedersen (2005)).

[4]Since Kraus and Stoll(1972), research has documented the permanent price impact of block trades. Bertsimas and Lo (1998) show that dividing up a large order evenly throughout the day optimally minimizes execution costs. Similarly, Seppi(2000) derives an equilibrium where a strategic informed institution "breaks up" a large trade into a series of smaller trades.

(which has become larger), what has not changed as much is the total volume contained within a sequence of consecutive buy or sell transactions. Post-2000, the average size of a string of consecutive transactions has remained stable at approximately 5,500 shares, and the standard deviation has remained stable at approximately 20,000 shares. Therefore, if we use sequences of consecutive transactions, in which we collapse any consecutive buy or sell transactions into a single observation by adding over volume, to study price impact, there are two advantages. One advantage is that there is now significant variation in the size of a trade. The other is that the dataset is well-behaved throughout time, unlike individual transactions, which have increased exponentially and decreased in size over time. Using a collapsed dataset also has the advantage of making the dataset far less cumbersome.

After documenting these changes in trading behavior, the first objective of our paper is to document the shortcomings of popular empirical market microstructure models when tested against current data. Specifically, we examine four models: first, Glosten and Harris (1988), which presents a model in which price changes depend on the the size of a trade. Second, Sadka (2006), which is similar to Glosten and Harris except that price changes depend on the information revealed by the *unanticipated* size of a trade. We consider it particularly important to examine the Sadka model given that the time series estimates of the model's parameters for the market portfolio are made available on Wharton Research Data Services (WRDS) and are widely used. Third, Huang and Stoll (1997), which presents a model in which additional costs are contained in the trade itself, regardless of the quantity, due to the market maker deviating from his favored zero inventory position; given the small size of individual trades in current data, it is questionable whether an order of 100 shares would cause any movement in the limit order book. Finally, we examine the model in Madhavan et al. (1997), which allows for autocorrelation in trade direction, but does not account for the size of the trade. All of these models account for the fixed cost component of transacting in a minimum number of shares (the bid-ask spread) by examining price changes due to moving from a buyer-initiated trade to a seller-initiated trade.

This first objective alone is important: while these models are widely used to provide estimates of market liquidity, particularly in equities, there is little in the literature evaluating their estimated liquidity measures compared with actual liquidity in the market. In this paper, we make an attempt to fill this gap in the literature. In particular, we evaluate these models based on how well their implied effective spreads compare to the actual effective spreads observed in the market. We would expect to see underestimation of the true effective spread, especially in the later portion of our sample period, because the data do not allow for proper identification of the variable cost component of trading due to decreased variation in the size of a transaction. As far as we know, ours is the first paper that calculates the effective spread implied from popular microstructure models, and compares it to the observed effective spread. The closest studies are Fialkowski and Petersen (1994) and Bessembinder and Kaufman (1997), both of which compare the posted bid-ask spread to the realized effective spread, and find that the former consistently overestimates the latter. We also evaluate these models based on the assumed random component of price changes - particularly, we should expect to see zero serial correlation, based on the assumptions common across the empirical market microstructure models discussed above.

We use SPY trade and quote data for the period January 1997 to November 2009, for a total of 3,272 trading days. SPY is the most liquid of the ETFs with a current daily trading volume of approximately 1.5 million shares, and the underlying 500 stocks constitute on average about 70% of the market capitalization of all exchange traded stocks in the USA (in 2008). We have two reasons for focusing on the SPY ETF: one, because we want to see how popular empirical microstructure models perform against high-frequency data, and the SPY is one of the most frequently-traded securities; and two, because the SPY, which provides a proxy for the market portfolio and is thus priced mostly by public information, would have few price changes due to adverse selection, allowing us to emphasize the role that order-splitting has in determining price changes.

We find that all four models consistently underestimate the true effective bid-ask spread: while the spread is only underestimated by about 7 percent in 1997-2006, the underestimation rises sharply to 41 percent in 2007-2009. We conjecture that this increase is due to changes in trading patterns: in 1997-2006, the average size of a trade is 2,700 shares (with a standard deviation of 15,000 shares), while in 2007-2009, it is only 400 shares (with a standard deviation of 6,600 shares). At the same time, the average number of consecutive buys or sells has increased from four to twelve. This corroborates our story that it is becoming increasingly common for traders to split their large orders into many smaller orders, in order to minimize price impact. The spread is underestimated because empirical market microstructure models only identify the price impact of a single trade, when we should be identifying the cumulative price impact of many small trades, as there is not enough variation in the size of a single transaction. All four models also model their residual price changes as white noise processes. We examine these residual price changes and find that there is still a significant degree of negative autocorrelation throughout the sample period, thereby identifying a second problem in these models. We conjecture that the splitting of large orders into smaller orders may also be a cause of this serial correlation.

The second objective of our paper is to then modify the transactions dataset to attenuate the problem of order splitting. Specifically, we aggregate consecutive buy or sell orders into a single trade, summing over the volume within those consecutive trades, and taking the size-weighted price. By aggregating over consecutive buys or sells, we observe the average price we pay for a total share quantity. Thus, we can identify both the costs due to trading in large volumes, and the order-processing costs due to switching between small buyer-initiated to seller-initiated trades. A Glosten-Harris type model based on these data produces effective spread estimates for the SPY that now underestimate the spread by only 1 percent in 1997-2006, and only 18 percent in 2007-2009. The negative autocorrelation in price changes is also reduced.

With a little bit of introspection, our findings should not come as a major surprise. These models were developed before electronic, order-driven markets became more prevalent, as predicted by Glosten (1994). Information contained in trades nowadays is processed very quickly, sometimes with even faster reversion in prices if the trade is detected to have no information component. Also, a centralized market-maker continues to play less of a role in the stock market, as order-driven markets become more popular and every trader effectively becomes a market-maker. Hence the price discovery mechanism may have changed with the popularity of electronic trading, and the reduction in information gathering costs due to web-based information becoming available to investors.

This paper is organized as follows. Section 2 describes the data used in the analysis. Section 3 documents the changes in trading behavior. Section 4 describes the four empirical market microstructure models used to model high-frequency price changes and the results from these models, and how they fail to properly capture the effective bid-ask spread. Section 5 presents a model in which sequences of consecutive buys or sell are aggregated into a single observation, and shows, via simulation and empirical evidence, that this helps to correct the problem of effective spread underestimation. Section 6 concludes.

2. Data

Our study focuses on the S&P 500 exchange-traded fund, SPY. We choose this security because it measures returns on a portfolio of securities that represent approximately 70 percent of total market capitalization, and thus we believe it is important to analyze, given its representation of the market. It is also among the highest-traded securities on a daily basis, and is frequently used by professionals for hedging and forming zero-cost arbitrage portfolios. As a robustness check, we also examine the Dow Jones exchange-traded fund, DIA.

SPY and DIA data were obtained from the NYSE Trade and Quote (TAQ) database for the period 1997 to 2009. We require the trade data to satisfy some technical conditions[5], and that the trades take place within normal business hours. We require the quote data to also satisfy some technical conditions[6], take place within normal business hours, for the bid price, ask price, bid volume, and ask volume to be greater than zero, the ask price to be greater than the bid price, and for the bid price to be within 10 percent of the ask price. We only use National Best Bid and Offer (NBBO) quotes on the second, which we determine using the NBBO algorithm provided by Wharton Research Data Services (WRDS).[7]

We match trades to contemporaneous quotes, and if there is no contemporaneous quote, we use the quote that is most recent, as suggested by Ellis et al. (2000). Other papers suggest matching trades to quotes lagged one second or greater - since we are dealing with securities that trade at a very high frequency, we believe it is reasonable that most trades are based on quotes displayed on that second, or the closest previous second otherwise.

With trades matched to quotes, we then determine whether a trade is buyer- or seller-initiated. We use the Lee-Ready (1991) algorithm - if the transaction price is above the bid-ask midpoint, then the trade is buyer-initiated ($D_t = +1$), and if the trade is below the bid-ask midpoint, then the trade is seller-initiated ($D_t = -1$). If the trade takes place at the midpoint, then a trade is classified as a buy or sell using the tick-test: it is a buy if the most recent price change was positive, and a sell if the most

[5]Specifically, for trades we need corr in $(0, 1, 2)$ (trades that are not cancelled or corrected) and cond not in (“O", “Z", “B", “T", “L", “G", “W", “J", “K") (trades without unusual conditions).

[6]That is, we require mode not in $(4, 7, 9, 11, 13, 14, 15, 19, 20, 27, 28)$ (BBO-eligible quotes).

[7]Hasbrouck (2010) also has a calculation of the NBBO which provides the NBBO tick-by-tick, but we do not use because we only require NBBO quotes on the second.

recent price change was negative.[8] Other papers use different classification schemes: Ellis et. al (2000), for example, uses the tick test for all trades, except for those that occur at the bid or at the ask.

3. Changes in Trading Behavior

Figure 1 displays histograms of the size of individual trades for the SPY for three periods: 1997 to 1999, 2000 to 2006, and 2007 to 2009. Pre-2000, we see that about 45 percent of individual trades are 500 shares less, 15 percent are greater than 500 shares and less than or equal to 1,000 shares, while the remaining 40 percent of trades are greater than 1,000 shares. In 2000 to 2006, about 75 percent of transactions are less than or equal to 500 shares, while in 2007 to 2009, about 85 percent of transactions are of this size. We also see that, in the later period, there are very few transactions of more than 2,000 shares ("sizebins" four to ten). The normal kernel density estimators, which overlap each of these histograms, make it clear that more transactions are concentrated in the lowest-sized bin in the latest sample period. The increasing popularity of smaller-sized transactions is likely because of lower bid-ask spreads, increasingly sophisticated technology that allows for more accessibility by individual traders, and institutional traders that divide their large transactions into smaller transactions in order to minimize price impact.

Table 1 provides more information. We see that, for the SPY, the average number of seconds between trades has steadily decreased over time, from 67.5 seconds in 1997 to 0.1 seconds in 2009. This decrease is likely because of the sheer increase in the number of transactions over time: on average, there were about 200 buy and sell transactions, each, in 1997, while there were about 250,000 in 2009. The average volume, in shares, has increased from 400,000 shares to 130.0 million shares. We see similar patterns for the Dow Jones ETF (DIA).

Expectedly, the average size of an individual transaction has substantially and steadily decreased over time - the average size of a buy or sell order in 1997 is 5,600 shares, while in 2009, it is only 400 shares. However, over this same time period, the average number of consecutive buy or sell transactions has significantly increased, from 2.3 consecutive buy or sell orders in 1997 to 11.9 in 2009 (see the "durtrade"

[8] Our only special case is when we calculate the parameters of the Madhavan, Richardson and Roomans model (which is discussed in the next section) - we only classify trades that take place at the bid or at the ask - all other trades are classified as neither ($D_t = 0$).

column). This provides evidence that traders may be splitting up their large orders into many smaller orders. Interestingly, the average volume contained in a sequence of consecutive buys or sells ("Dursize") has remained relatively stable - from 2001 to 2009, for example, we see that the average volume stays at around 5,500 shares. This is not the case pre-2001 - however, this may be the result of aggregating over consecutive buys or sells that are too many seconds apart. The average time elapsed within a sequence of consecutive buys or sells is very high pre-2001, as evidenced by the "Durtime" column. Figure 2 plots the monthly time series averages of both the size of a transaction and the number of transactions contained in a sequence of buys or sells, for both SPY and DIA - it is clear that the size of a transaction has been steadily decreasing at the same time the number of trades contained in a sequence of buys or sells has been increasing. Figure 3 makes it clear that the average total volume contained in a sequence of consecutive buys or sells has remained relatively stable.

We emphasize the stability of total volume contained in sequences of consecutive buy or sell transactions by referring to Figure 4. This figure displays histograms of the total volume contained in sequences of buy or sell transactions for the same three periods: 1997 to 1999, 2000 to 2006, and 2007 to 2009. While the proportion of volume in the smallest share bin still slightly increases as we move into the latest sample (only by about 5 percent), we clearly see that the histograms are quite similar across the three sample periods. The normal kernel density estimators in each histogram emphasize the similarities across these histograms.

A problem with using transaction size to identify the variable cost of a transaction is that when there is little variation in transaction size, we have no identification. Table 2 reports that the variation in the size of individual transactions has significantly decreased over time - from 23,000 shares in 1997 to 5,100 shares in 2009. However, the variation in the size contained in a sequence of consecutive buys or sells has remained relatively constant at 20,000 shares for the 2001 to 2009 period. Pre-2001, the deviation is higher, but again, this may be the result of aggregating consecutive buys or sells that are too far apart in time - we see from this table, and Table 1, that both the mean and standard deviation of time between transactions is fairly high pre-2001.

4. Testing the Older Models

4.1. Methodology

Our first objective is to test the efficacy of empirical market microstructure models using current, high-frequency data. We will test these models by calculating the effective bid-ask spread implied by each model's estimated parameters, and comparing it to the true estimate of the effective bid-ask spread. The true effective spread is calculated as twice the absolute difference of the price of a transaction p_t and its corresponding bid-ask midpoint m_{t-1} (which we will assume is reflective of the true price process), where the price and midpoint are matched using the Lee-Ready algorithm. We calculate the daily effective spread by averaging over the individual effective spreads in that day (*Avg* represents the "Average" operator: the sum of all values divided by the number of observations). Thus our true measure of the daily effective spread is given by:

$$ES_d^{true} = 2Avg(|p_t - m_{t-1}|).$$

Next, we focus on several empirical market microstructure models and calculate their implied effective bid-ask spreads, so that we can compare these implied effective spreads to the true effective spread.

Each model assumes that the true, unobserved price process m at trade t is a function of either signed trade volume (DV_t: positive for buys, negative for sells), a buy/sell indicator (D_t: +1 for a buy, -1 for a sell), or both, plus some random white noise component:

$$\Delta m_t = f(D_t, DV_t) + \epsilon_t. \tag{1}$$

That is, each model assumes that at least a part of trading volume and/or the buy/sell indicator gets incorporated into the true price process. The observed price process, which we obtain from the TAQ trade files, is also a function of the signed trade volume or the trade indicator:

$$p_t = m_t + c_0 D_t + c_1 DV_t. \tag{2}$$

That is, buyers pay in excess of the true price because of the transitory costs of trading, and sellers receive less.

The differences in the empirical models will mostly come in different functional forms f, and assumptions on c_0 and c_1. Combining the two equations above, by taking differences in the equation (2) and substituting in equation (1), provides us with a testable empirical model:

$$\Delta p_t = f(D_t, DV_t) + c_0 \Delta D_t + c_1 \Delta DV_t + \epsilon_t. \tag{3}$$

Briefly, we now outline the specifics of the four empirical microstructure models that we will test:

- Glosten and Harris (1988) (GH): Assumes $f(D_t, DV_t) = z_0 D_t + z_1 DV_t$. They also incorporate rounding of the price process due to discreteness of the price grid into the error term. We will not incorporate this discreteness because the small tick size ($0.01) in most of our sample period makes this an insignificant concern.
- Sadka (2006): Assumes $f(D_t, DV_t) = z_0 UD_t + z_1 UDV_t$, where UDV_t is the residual component from an AR(5) on DV_t, and UD_t, the unexpected component of trade, is calculated using UDV_t (see Sadka (2006)); the intuition is that the unexpected component of demand is what gets incorporated into the true price process. This model is similar to GH.
- Huang and Stoll (1997) (HS): Assumes $f(D_t, DV_t) = z_0 D_{t-1}$ and $c_1 = 0$. They assume that once a market maker makes a trade, he adjusts the bid-ask midpoint to encourage the next trade to be in the opposite direction, in order to balance his inventory. Volume does not factor into this model.
- Madhavan, Richardson and Roomans (1997) (MRR): Assumes $f(D_t, DV_t) = z_0 UD_t$, where UD_t is the residual component from an AR(1) on D_t. They incorporate first-order autocorrelation ρ into D_t, and have a slightly different definition of D_t in that it can equal zero if a transaction takes place between the corresponding bid and ask prices. With no assumptions on their model's ϵ_t, they estimate their model using GMM.

Note that c_0 and c_1 represent the transitory components of price changes, while z_0 and z_1 represent the permanent components of price changes.

With daily parameter estimates obtained from equation (3), we can now calculate the effective bid-ask spread implied by each of the four models. We define the estimates of the effective bid-ask spread implied by each model as twice the absolute difference of the transaction price minus the previous bid-ask midpoint, net of the noise from ϵ_t. This measure, which is a function of the size of the transaction, captures the round-trip cost of buying and then immediately selling a certain number of shares, is calculated per day (d):

$$ES_d^{implied} = 2Avg(|p_t - m_{t-1} - \epsilon_t|), \tag{4}$$

where we assume that, on average, the bid-ask midpoint reflects the true price of the security. Using the structural equations from each of the four models, we get the following measures of the implied effective bid-ask spread (V_t represents the size of the trade):

$$\begin{aligned} GH: &\quad ES_d^{GH} = 2Avg(|c_0 + z_0 + (c_1 + z_1)V_t|) \\ Sadka: &\quad ES_d^{Sadka} = 2Avg(|c_0 D_t + c_1 DV_t + z_0 UD_t + z_1 UDV_t|) \\ HS: &\quad ES_d^{HS} = 2Avg(c_0) \\ MRR: &\quad ES_d^{MRR} = 2Avg(|c_0 D_t + z_0(D_t - \rho D_{t-1})|). \end{aligned}$$

The calculation of the effective spread for the Huang-Stoll model slightly differs in that we take the difference of the observed price and the contemporaneous true value, because contemporaneous trades are not incorporated into the true price process.

4.2. Results

Figure 5 presents the monthly time-series estimates of the effective bid-ask spreads for the SPY from each of the four microstructure models. All four models are plotted against the realized effective bid-ask spread. We see that underestimation of the effective spread is consistent across time, across all four models, although the underestimation is fairly small in the 1997-2006 period - on average, the estimated spread underestimates the realized spread by about 7 percent. However, in 2007-2009, the spread is underestimated by 41 percent. Clearly, there is some structural shift that is driving this sharp increase in underestimation - we conjecture that it is because splitting large orders into very small orders has become increasingly common, and that the models are not identifying the cumulative impact that many small orders may have on price changes. We also check the effective bid-ask spreads for the DIA, and find similar results, as shown in Figure 6. In 1997-2006, the effective spread is underestimated by 6 percent, while in 2007-2009, it is underestimated by 46 percent.

As an additional test, we also plot the residual serial correlations from the GH model versus the serial correlations in observed tick-by-tick price changes. Henceforth, we only focus on the GH model, as the remaining models produce similar results (with the exception of the MRR model, which always performs slightly worse). Under the assumptions of these models, while price changes should exhibit negative serial correlation due to bid-ask bounce, the price change residuals should exhibit zero serial correlation, as the bid-ask bounce effect has been removed from the price changes. However, we see from Figure 7 that the residual serial correlation is still significantly negative, averaging about -0.26. However, in 1997-2006, this average is -0.27, while in 2007-2009, it is -0.22, suggesting that the shift in trading patterns may not be influencing the negative autocorrelation. Nonetheless, this is still a significant problem in these models. We conjecture that the effects of trading volume have not been fully incorporated into price changes, as these models only consider the size of individual trades, and not the effect that many consecutive smaller trades can have on prices.

5. Adapting to High Frequency Data

We have examined several empirical market microstructure models and have shown how they underestimate the realized effective spread, especially for the post-2006 period. We stress that the spread is underestimated because there is little variability in trading size in contemporary data - with no variability, we have no identification. Next, we discuss how we will modify contemporary data so that we produce more accurate measures of the realized effective spread.

We modify the transactions dataset for the SPY by aggregating sequences of consecutive buys or sells into a single observation, summing over volume and taking the size-weighted average price. This way, we can identify the average additional price we pay for transacting in a certain number of shares. The original microstructure models may underestimate the variable cost component due to trade size because most small trades do not have any price impact at all - it is the accumulation of small trades that have a significant price impact.

This simple aggregation is similar to an aggregation done in Hasbrouck (1991) - there, he assumes that trades that occur within five seconds of each other, with no intervening quotes, are accumulated into a single observation, which is done to alleviate the problem of reporting fragmentation. The issue in our paper is also the fragmentation of trades, with the main difference being that we have many intervening quotes between transactions. Sadka (2006) also alleviates reporting fragmentation by aggregating trades that occur at the same price within the same second - we aggregate trades that occur across seconds at different prices, as it is very possible that a trader executing many small trades will have to pay higher prices as he goes deeper into the limit order book, and tries to minimize the price impact of his trade by fragmenting his orders.

5.1. Simulation Results

We first show, via simulation, that it can be the case that the effective bid-ask spread will be underestimated if we do not aggregate small trades. Appendix A and the accompanying Figure 8 show that the spread is underestimated in the traditional Glosten-Harris model because the marginal cost of trading is increasing in the cumulative trading size. If we assumed that all trades are very small, which is the case in the latter portion of our sample period, then the Glosten-Harris model is unable to identify the increasing marginal costs incurred by subsequent trades. However, by summing over all volume contained within a sequence of consecutive buy or sell orders, the Glosten-Harris model is able to identify a the price impact of high volume trades.

The simulation results, reported in Figure 8, show that the effective spread is underestimated when individual trades are used to identify the price impact of a trade. This underestimation is increasing in the "beta" of the security, where beta represents the true price impact of a transaction. The underestimation of the effective spread becomes more pronounced as beta increases because single trades alone are unable to identify the variable costs of trading, due to lack of variation in the single trades (in this simulation, we assume every trade is of size one). However, when we aggregate over consecutive buys and sells, the underestimation of the effective spread is nowhere near as severe because we now have enough variation in trade size.

5.2. Model and Empirical Results

Next, we take the actual dataset for the SPY, for the period 1997 to 2009, collapse sequences of consecutive buys or sells into single transactions, and see how the Glosten-Harris model performs using this dataset. Let s index the s-th string of consecutive buys or sells in the series of trades t. Then we have the following definitions:

$$V_s = \sum_{t \in s} V_t,$$

$$p_s = \sum_{t \in s} \frac{V_t p_t}{V_s},$$

where V_t is the size of an individual trade t, V_s is the total size within a sequence of consecutive buy or sell transactions, p_t is the price of an individual trade, and p_s is the average price paid within a sequence of consecutive buy or sells transactions. With this new dataset defined, we can now examine the dynamics of size-weighted price changes, in response to changing between buy and sell orders, and the total size within that order. We have the following testable equation:

$$\Delta p_s = \alpha \Delta D_s + \lambda V_s + \epsilon_s,$$

where D represents the buy/sell indicator, and ϵ_s is the residual average price change, which is assumed to be zero mean i.i.d. By construction of the dataset, D_a will always alternate between -1 and 1.

Figure 9 reports our results. Clearly, the estimated effective spread closely matches the realized effective spread throughout the sample period, although there is still some slight underestimation in 2007-2009. For the SPY, we now have that the estimated effective spread underestimates the realized effective spread by only 1 percent in 1997-2006, and 18 percent in 2007-2009 (similar results hold for DIA). While the underestimation is still very much there in the latter period, we believe that by aggregating consecutive buy or sell orders, the Glosten-Harris model is better-able to pick up on price changes because of better identification via higher trade size variability. Also, Figure 10 reports the serial correlation in the residuals from our model. In 1997-2006, the serial correlation is now -0.17, while in 2007-2009, it is now -0.13. By aggregating the data in consecutive buy or sell orders, we have reduced both underestimation error in the effective bid-ask spread, and negative serial correlation in the residual component of price changes.

6. Conclusion

We test for potential misspecification in four well-known market microstructure models – Glosten and Harris (1988), Sadka (2006), Huang and Stoll (1997), and Madhavan, Richardson, and Roomans (1997) – by calculating the daily effective spread implied by the parameters in each of the models and comparing them to the observed daily effective spread using SPY data for the period 1997 to 2009. We find that, on average, the estimated effective spread underestimates the actual effective spread by about 7 percent in 1997-2006, and 41 percent in 2007-2009. The models also produce price change residuals that have significant negative autocorrelation, contradicting the assumptions of those models.

We conjecture that the errors in estimation are caused by changes in trading patterns. In 1997-2006, the average size of a single trade was 2,700 shares, with a standard deviation of 15,000, while in 2007-2009, it was only 400 shares, with a standard deviation of 6,600. With little variation in trade size in the late sample period, it is difficult to identify price changes due to trade size. However, at the same time, the average number of consecutive buys or sells has increased from four to twelve, which leads us to believe that traders may be trading in smaller quantities because they are splitting up their large trades into several smaller component, likely to minimize the price impact of their large trade. If we collapse consecutive buy or sell orders into a single transaction by summing over volume, we find that little has changed from the early to late sample periods - the size, variation and distribution of these collapsed trades stays somewhat stable over time.

We show, first by simulation and then using real data, that when we modify the transactions dataset by collapsing sequences of consecutive buys or sells into a single transaction, the Glosten-Harris model estimates the true costs of trading quite well. Underestimation of the effective bid-ask spread is now only 1 percent and 18 percent for the 1997-2006 and 2007-2009 period, respectively. We also find some success in reducing the negative serial correlation in the price change residual component. We conclude that by aggregating strings of consecutive buys or sells, due to the fact that it has become increasingly common for traders to split up their large orders, we improve the ability to model the price change process and produce accurate measures of the effective bid-ask spread.

Appendix A

In this section, we describe the details of the simulation to show how the effective spread is underestimated when traders split their large orders into many smaller orders. We need to model trading costs so that they are increasing in the cumulative order size contained within a string of consecutive buys or sells. Our model is as follows. First, we assume that the fundamental value of the security m follows a random walk process:

$$m_t = m_{t-1} + \varepsilon_t,$$

where ε_t is a white noise process. Next, we assume that the price process evolves as follows:

$$p_t = m_t + D_t \left(\alpha + \beta \log \left(\sum_{i=0}^{j} I \left(D_{t-i} = D_t \right) \right) \right).$$

That is, we assume that the price p_t is equal to the fundamental value plus the spread component due to order-processing costs α, plus another spread component due to the size of the trade. We assume that the size of each trade is equal and small, so that the total size of the trade is simply the sum of the small trades - this is akin to a trader breaking up his large order into many small, equal-sized orders. The price is a function of the natural log of the total size of the trade because we assume that total trading costs are convex in the size of the trade. β represents the cost paid on the log of total trades.

In each simulation, we generate the price and trade sign process as follows. We assume that the number of strings of consecutive buys or sells in a single day is 1,000. In each string, we assume that the number of trades can be between 6 and 15 with equal probability, and that the first string is buyer-initiated, and alternate between buyer-initiated and seller-initiated strings. The m_t process is generated from a random draw of $\varepsilon_t \sim N(0, 1)$. Using the generated m_t and D_t processes, we can simulate p_t.

Then, we run an OLS regression using the following model:

$$\Delta p_t = c_0 \Delta D_t + z_0 D_t + \epsilon_t, \tag{5}$$

and calculate the implied effective spread as $ES_{imp} = Avg \left| c_0 D_t + z_0 D_t \right|$.

Then, we modify the data as suggested in the paper. Denote s as the s-th string of consecutive buy or sell orders. We have the following definitions:

$$\begin{aligned} D_s &= \text{trade sign for string } s, \\ DV_s &= \text{number of trades in a string } s, and \\ p_s &= \frac{\sum_{t\in s} p_t}{DV_s}. \end{aligned}$$

We then run an OLS regression using the following model:

$$\Delta p_s = c_0 \Delta D_s + z_0 DV_s + \epsilon_s,$$

and calculate the modified implied effective spread as $ES_{mod} = Avg\,|c_0 D_s + z_0 DV_s|$.

Finally, the realized spread is calculated as

$$kES_{real} = Avg\,|p_t - m_t|\,.$$

For each level of β we perform the simulation 3000 times and plot the averages of ES_{mod}, ES_{real}, and ES_{imp}. See figure 8 for the plots.

References

[1] Bertsimas, Dimitris, and Andrew W. Lo, 1998, Optimal control of execution costs, *Journal of Financial Markets* 1, 1-50.

[2] Bessembinder, Hendrik, 2003, Trade Execution Costs and Market Quality after Decimalization, *Journal of Financial and Quantitative Analysis* 38, 747-777.

[3] Bessembinder, Hendrik, and Herbert M. Kaufman, 1997, A Comparison of Trade Execution Costs for NYSE and NASDAQ-Listed Stocks, *Journal of Financial and Quantitative Analysis* 32, 287-310.

[4] Brunnermeier, Markus K., and Lasse Heje Pedersen, 2005, Predatory Trading, *Journal of Finance*, 60, 1825-1863.

[5] Buti, Sabrina, Barbara Rindi and Ingrid M. Werner, 2010, Diving Into Dark Pools, Working Paper.

[6] Ellis, Katrina, Roni Michaely and Maureen O'Hara, 2000, The Accuracy of Trade Classification Rules: Evidence from Nasdaq, *Journal of Financial and Quantitative Analysis* 35, 529-551.

[7] Fialkowski, D. and Petersen, M.A., 1994, Posted versus effective spreads: good prices or bad quotes, *Journal of Financial Economics* 35, 269–292.

[8] Glosten, Lawrence R., 1994, Is the Electronic Open Limit Order Book Inevitable?, *Journal of Finance* 49, 1127-1161.

[9] Glosten, Lawrence R., and Lawrence E. Harris, 1988, Estimating the components of the bid/ask spread, *Journal of Financial Economics* 21, 123-142.

[10] Hasbrouck, Joel, 1991, The Summary Informativeness of Stock Trades: An Econometric Analysis, *The Review of Financial Studies* v4 n3.

[11] Joel, Hasbrouck, 2007, Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading, Oxford University Press.

[12] Joel, Hasbrouck, 2010, The Best Bid and Offer: A Short Note on Programs and Practices , Working Paper.

[13] Huang, Roger D., and Hans R. Stoll, 1997, The Components of the Bid-Ask Spread: A General Approach, *Review of Financial Studies* 10, 995-1034.

[14] Kraus, A., and H. R. Stoll, 1972, Price Impacts of Block Trading on the New York Stock Exchange, *Journal of Finance* 27, 569-588.

[15] Kyle, A. 1985, Continuous auctions and insider trading, *Econometrica*, 53, 1315-35.

[16] Lee, Charles M. C. and Mark J. Ready, 1991, Inferring Trade Direction from Intraday Data, *Journal of Finance* 46, 773-746.

[17] Madhavan, Ananth, Matthew Richardson, and Mark Roomans, 1997, Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks, *Review of Financial Studies* Vol.10 , 1035-1064.

[18] Ronnie, Sadka, 2006, Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk, *Journal of Financial Economics* 80, 309-349.

[19] Seppi, D., 2000, Equilibrium block trading and asymmetric information, *Journal of Finance* 45, 73-94.

[20] Stoll, H., 1978, The Supply of Dealer Services in Securities Markets, *Journal of Finance* 33, 1133-1151.

[21] Stoll, Hans R. and Christoph Schenzler, 2006, Trades outside the quotes: Reporting delay, trading option, or trade size?, *Journal of Financial Economics* 79, 615-653.

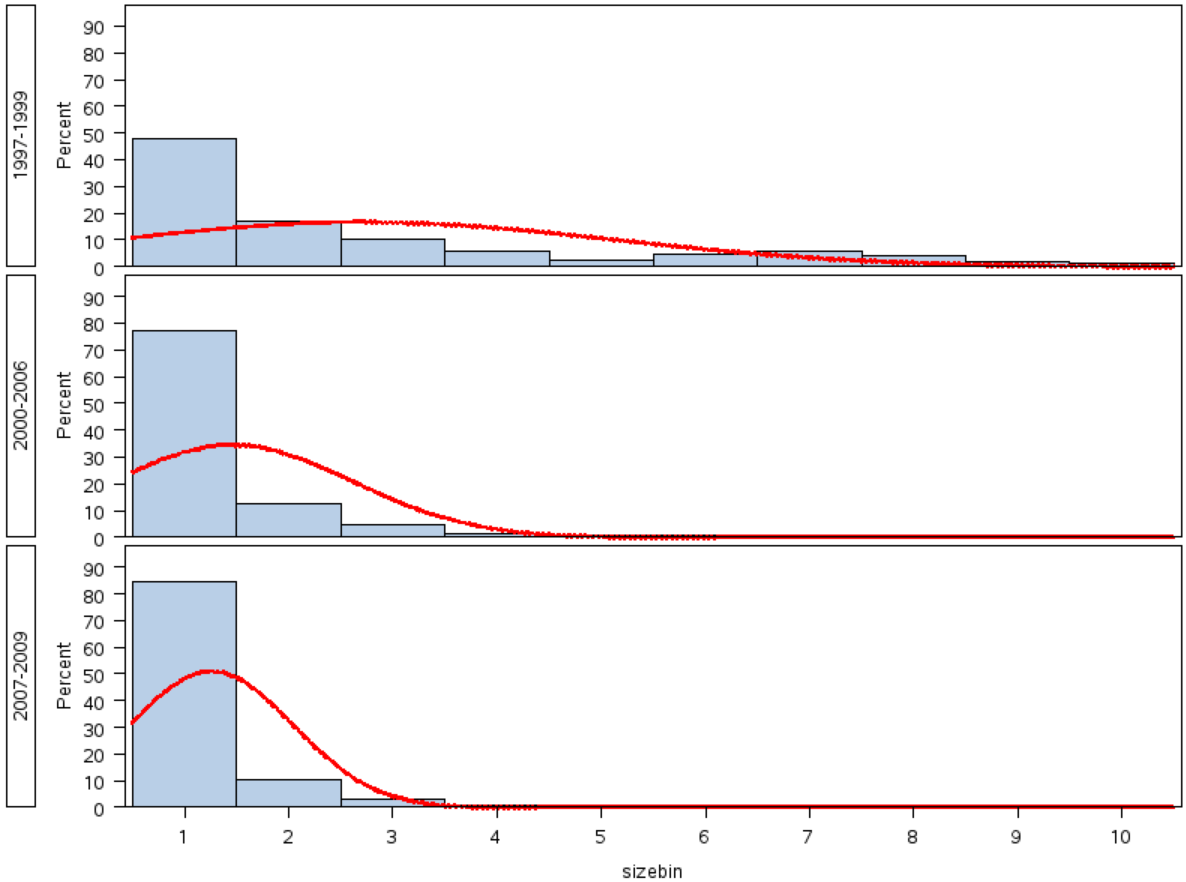


Figure 1: **Distribution of Trade Size.** This graph is a histogram of the size of individual trades for the SPY for three different eras: 1997-2000 (pre-decimalization), 2001-2006, and 2007-2009. The "size bins" on the horizontal axes are defined as follows: 1 (500 shares or less); 2 (501-1,000 shares); 3 (1,001-2,000 shares); 4 (2,001-3,000 shares); 5 (3,001-4,000 shares); 6 (4,001-5,000 shares); 7 (5,001-10,000 shares); 8 (10,001-25,000 shares); 9 (25,001-50,000 shares); and 10 (greater than 50,000 shares).

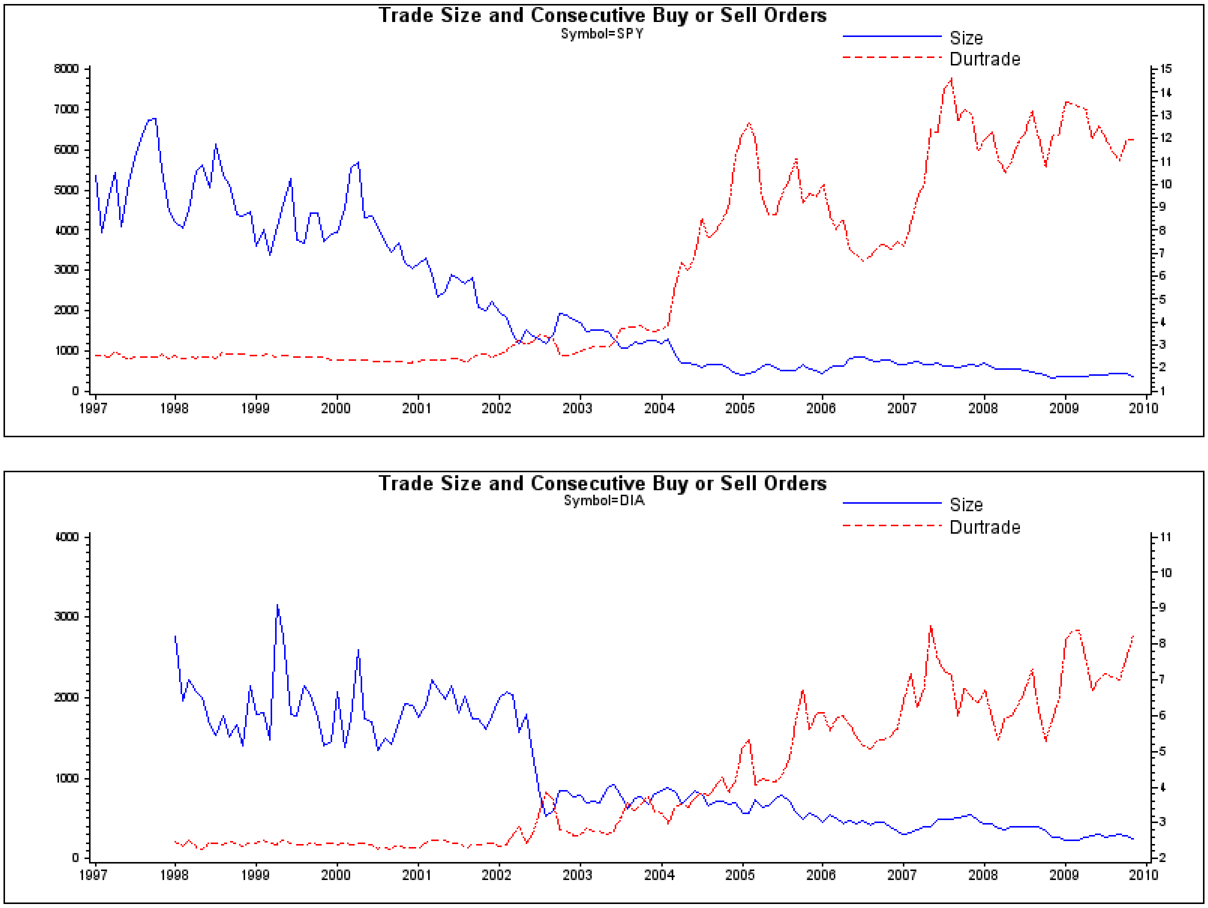


Figure 2: **Trade Size and Consecutive Buys or Sells.** The graphs plot the average size of a trade and the average number of trades in a string of consecutive buys or sells for the S&P500 ETF (SPY) and the Dow Jones ETF (DIA), respectively. The time period is 1997 to 2009 for the SPY, and 1998 to 2009 for the DIA.

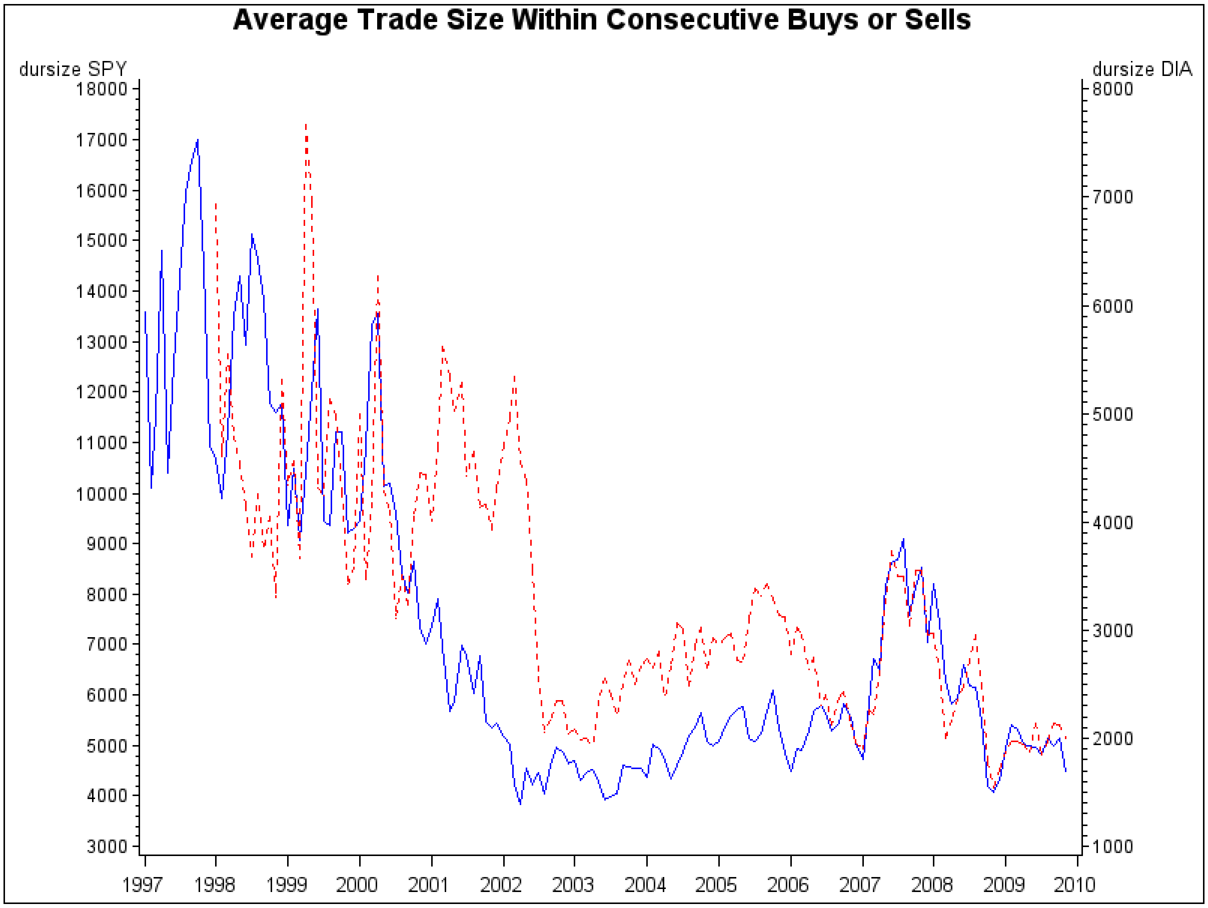


Figure 3: **Trade Size Within Consecutive Buys or Sells.** The graphs plot the average size within a string of consecutive buys or sells for the S&P500 ETF (SPY) and the Dow Jones ETF (DIA), respectively. The time period is 1997 to 2009 for the SPY, and 1998 to 2009 for the DIA.

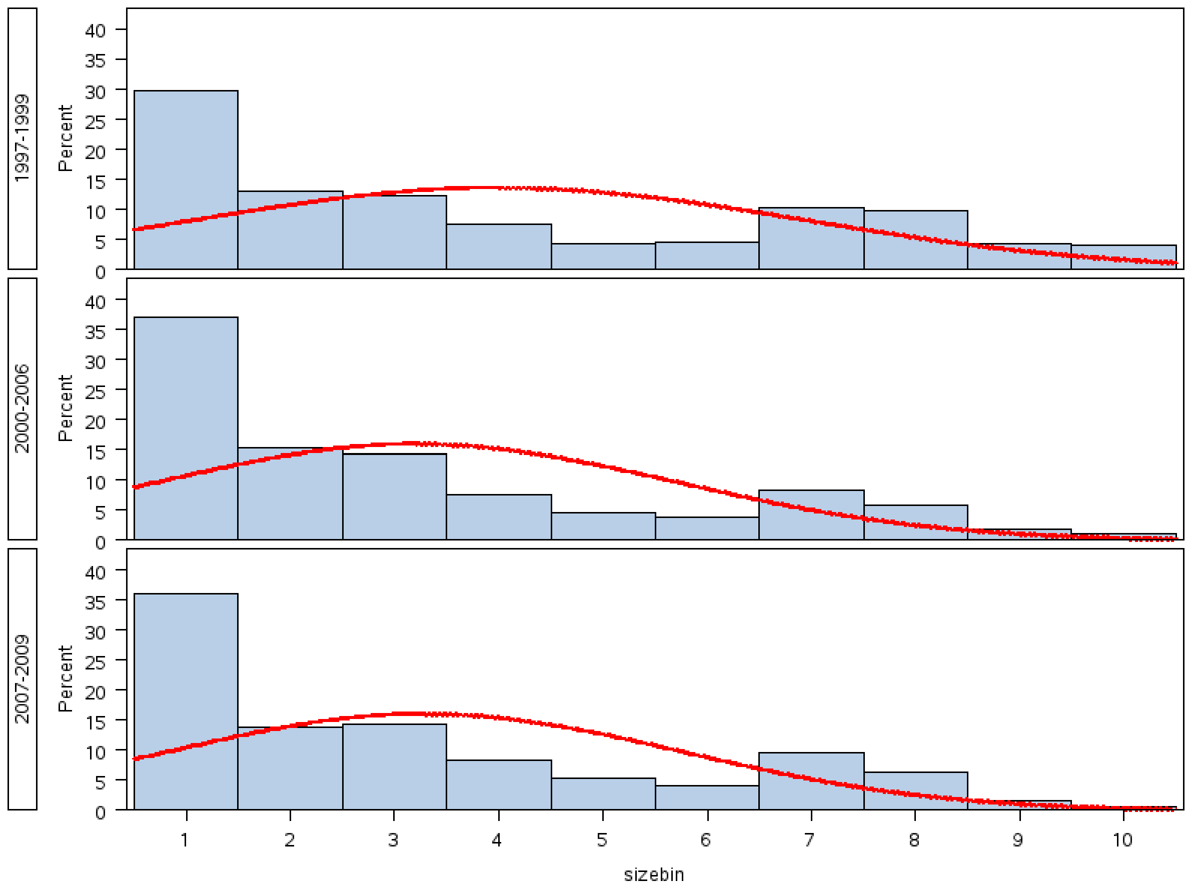


Figure 4: **Distribution of Trade Size.** This graph is a histogram of the total volume contained within sequences of consecutive buy or sell transactions for the SPY for three different eras: 1997-2000 (pre-decimalization), 2001-2006, and 2007-2009. The "size bins" on the horizontal axes are defined as follows: 1 (500 shares or less); 2 (501-1,000 shares); 3 (1,001-2,000 shares); 4 (2,001-3,000 shares); 5 (3,001-4,000 shares); 6 (4,001-5,000 shares); 7 (5,001-10,000 shares); 8 (10,001-25,000 shares); 9 (25,001-50,000 shares); and 10 (greater than 50,000 shares).

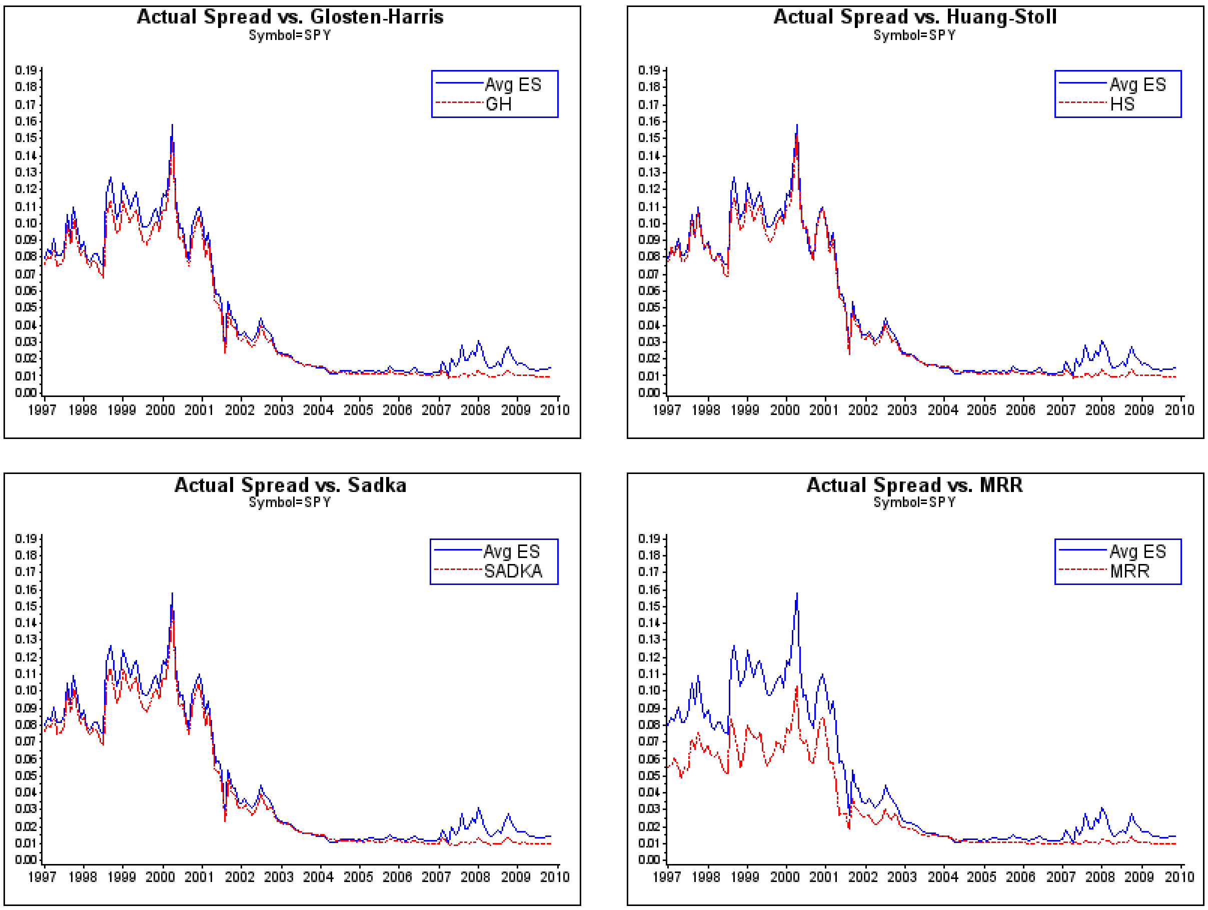


Figure 5: **Estimated versus Realized Effective Spreads.** All four graphs plot the mean monthly effective spread versus the estimated monthly effective spread for the S&P500 ETF (SPY). The four graphs estimate the monthly effective spread using the Glosten-Harris (GH), Sadka, Huang-Stoll (HS), and Madhavan-Richardson-Roomans (MRR) models, respectively, for the period 1997 to 2009. y-axis units are in dollars.

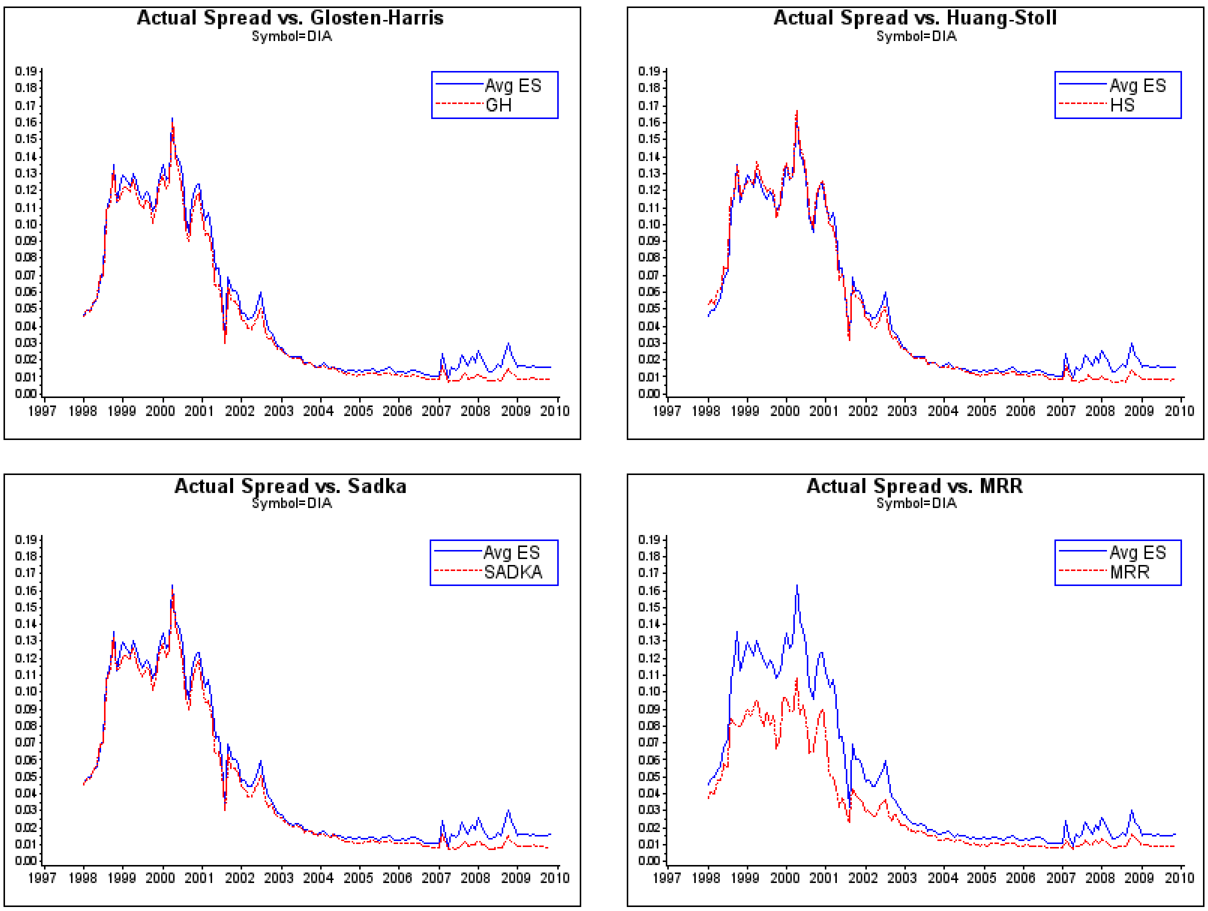


Figure 6: **Estimated versus Realized Effective Spreads.** All four graphs plot the mean monthly effective spread versus the estimated monthly effective spread for the Dow Jones ETF (DIA). The four graphs estimate the monthly effective spread using the Glosten-Harris (GH), Sadka, Huang-Stoll (HS), and Madhavan-Richardson-Roomans (MRR) models, respectively, for the period 1997 to 2009. y-axis units are in dollars.

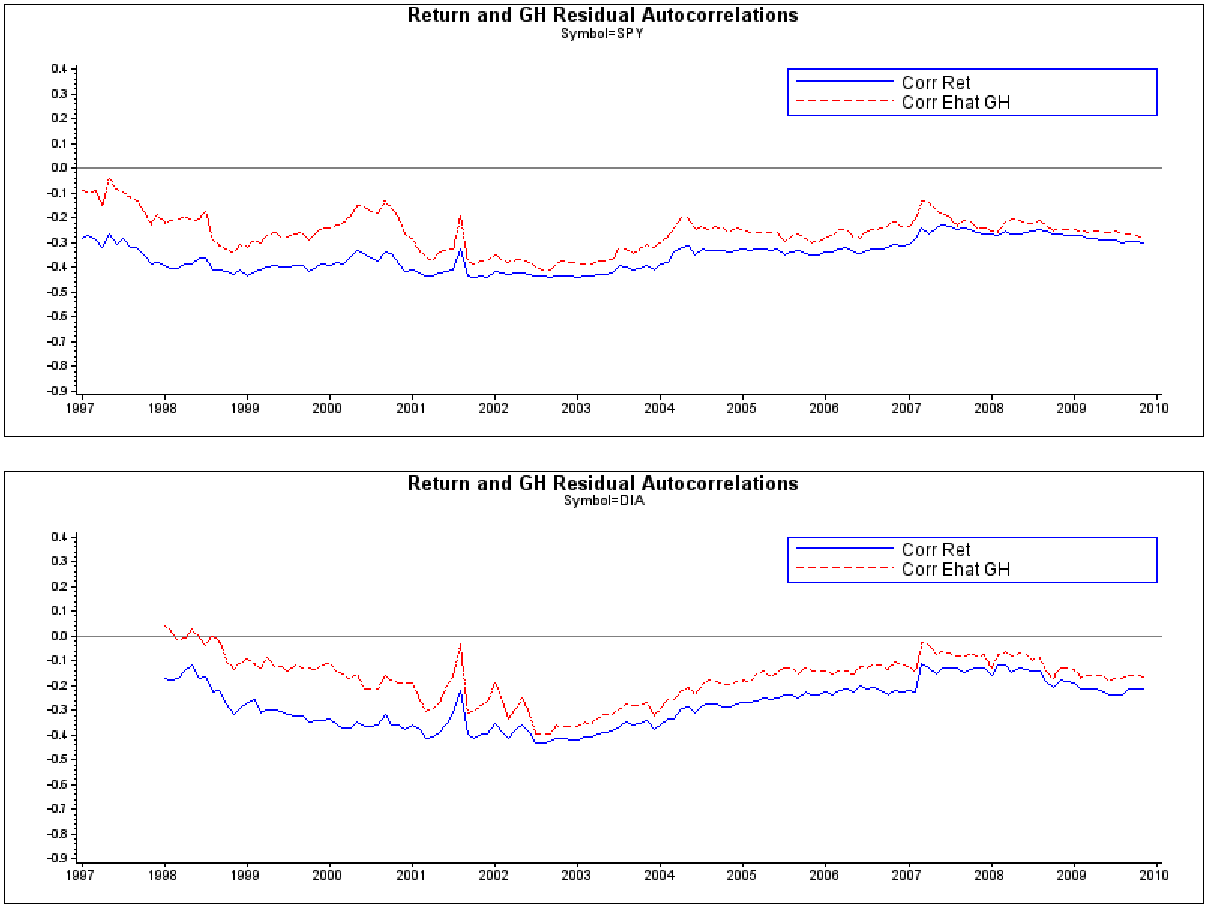


Figure 7: **Autocorrelations of Returns and Residuals.** The graphs plot the tick-by-tick autocorrelation in observed returns and autocorrelation in residuals from the Glosten-Harris model for the S&P500 ETF (SPY) and the Dow Jones ETF (DIA), respectively. The time period is 1997 to 2009 for the SPY, and 1998 to 2009 for the DIA.

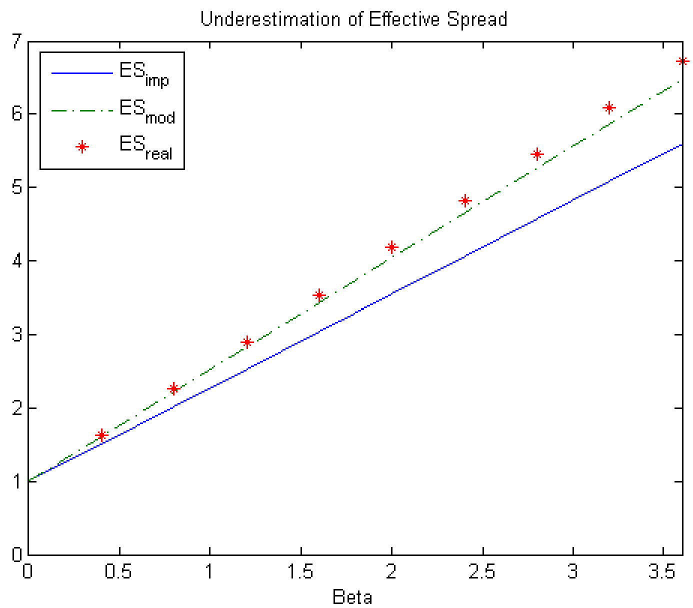
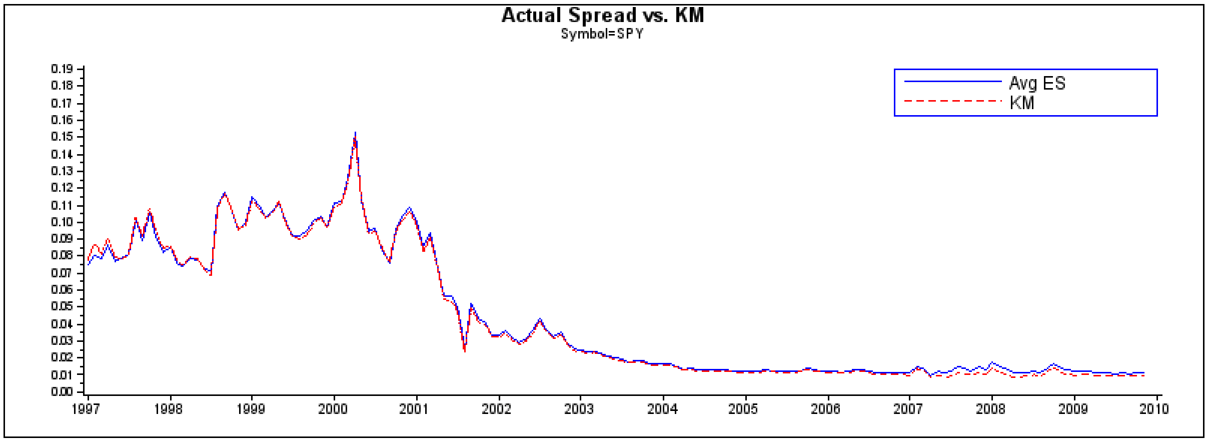


Figure 8: **Simulation of Effective Spreads.** This graph plots the simulated realized effective spread (ES_{real}), the effective spread implied by the Glosten-Harris using the simulated data (ES_{imp}), and the effective spread implied by the Glosten-Harris model from the simulated data, where transactions are aggregated over sequences of consecutive buys or sells by summing over volume and taking the size-weighted average price (ES_{mod}). The α (half-spread) coefficient is set equal to 1, and ε follows a $N(0,1)$ process (See Appendix A for more details).



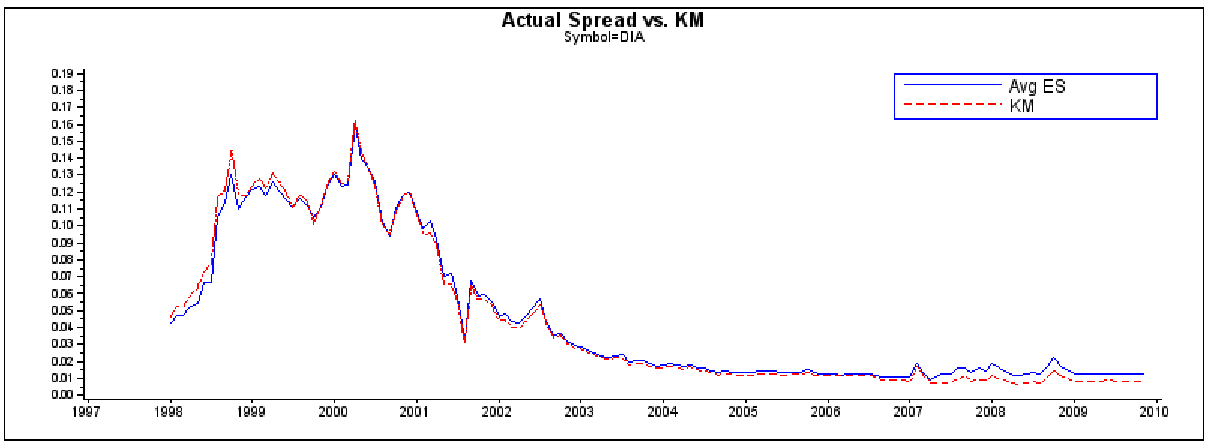
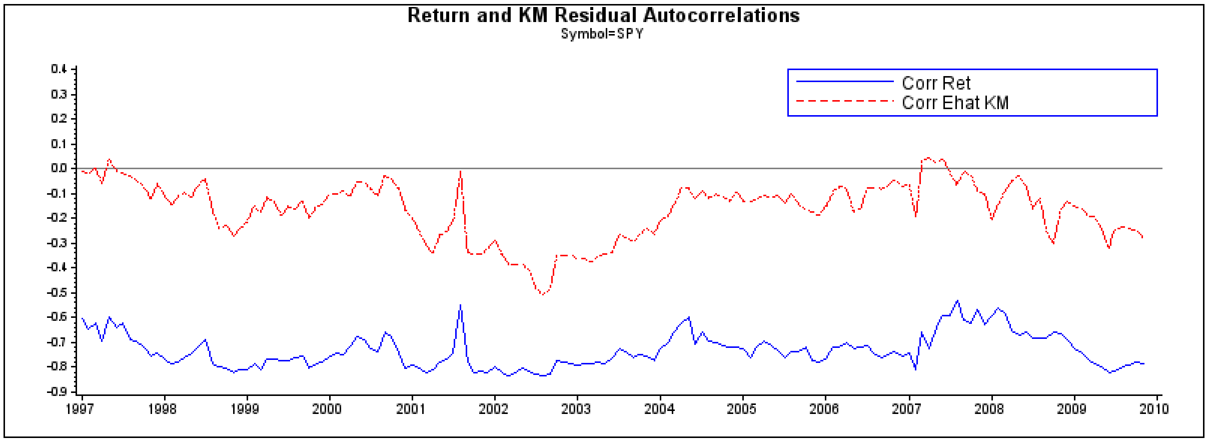


Figure 9: **Estimated versus Realized Effective Spreads.**The graphs plot the mean monthly effective spread versus the estimated monthly effective spread for the S&P500 ETF (SPY) and the Dow Jones ETF (DIA), respectively. The estimate of the monthly effective spread is based on the Kim-Murphy model (KM), which is based on mean price movements within strings of consecutive buys or sells. The time period is 1997 to 2009 for the SPY, and 1998 to 2009 for the DIA. y-axis units are in dollars.



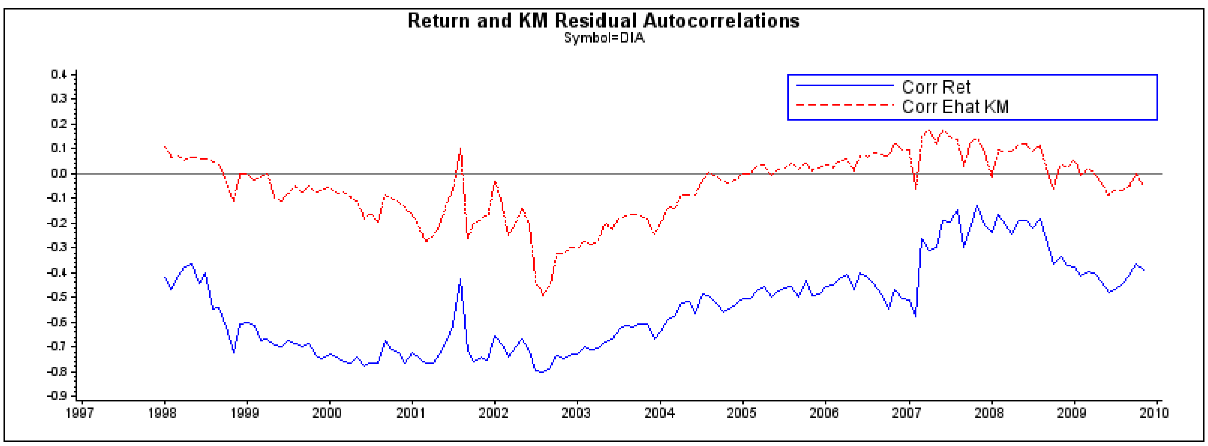


Figure 10: **Autocorrelations of Returns and Residuals.** The graphs plot the tick-by-tick autocorrelation in observed returns and autocorrelation in residuals from the Kim-Murphy (KM) model for the S&P500 ETF (SPY) and the Dow Jones ETF (DIA), respectively. The time period is 1997 to 2009 for the SPY, and 1998 to 2009 for the DIA.

Table 1: Panel A provides yearly summary statistics for the S&P500 ETF (SPY). Panel B provides yearly summary statistics for the Dow Jones ETF (DIA). The variables are defined as follows: "Buy" is the average number of buy orders per day; "Sell" is the number of sell orders per day; "Δtime" is the average number of seconds between trades; "Sizeday" is the average number of shares traded in a day, in millions; "Durtrade" is the average number of consecutive buys or sells; "Dursize" is the average number of shares traded within a string of consecutive buys or sells; "Durtime" is the average number of seconds elapsed within a string of consecutive buys or sells.

Panel A: SPY

Year	Buy	Sell	Δtime	Size	Sizeday	Durtrade	Dursize	Durtime
1997	222	192	67.5	5,605	2.4	2.3	12,990	80.5
1998	565	525	27.3	5,142	2.7	2.3	11,732	30.1
1999	639	589	20.2	4,277	5.3	2.2	9,348	22.1
2000	671	607	20.2	4,197	5.4	2.1	8,813	20.7
2001	1,889	1,829	8.1	2,738	9.7	2.1	5,655	8.2
2002	6,498	6,355	2.2	1,650	20.4	2.5	4,055	2.7
2003	8,542	8,197	1.5	1,384	23.0	3.1	4,227	2.4
2004	17,594	17,323	0.8	763	24.0	7.5	5,185	3.2
2005	35,134	34,862	0.4	546	37.9	10.6	5,665	2.4
2006	30,723	30,646	0.4	722	43.9	7.9	5,615	2.1
2007	82,485	83,046	0.2	660	108.6	11.3	7,433	1.3
2008	229,917	230,792	0.1	505	210.4	11.1	5,586	0.5
2009	250,704	249,047	0.1	400	1935.6	11.9	4,715	0.4

Panel B: DIA

Year	Buy	Sell	Δtime	Size	Sizeday	Durtrade	Dursize	Durtime
1998	130	113	118.6	1,894	0.4	2.3	4,341	138.3
1999	168	144	87.6	1,942	0.7	2.2	4,340	96.5
2000	307	268	49.7	1,753	1.6	2.1	3,776	51.9
2001	665	635	24.1	1,941	2.5	2.2	4,197	26.6
2002	2,875	2,853	8.5	1,317	5.2	2.5	3,085	9.8
2003	2,853	2,747	4.5	782	4.4	3.0	2,327	7.4
2004	2,805	2,726	4.7	784	4.3	4.0	3,106	11.0
2005	3,822	3,803	3.5	645	4.8	5.7	3,577	11.9
2006	5,698	5,637	2.3	456	5.1	6.5	2,981	9.7
2007	9,830	9,948	1.4	446	9.0	8.0	3,509	7.6
2008	22,617	22,775	0.7	376	15.9	6.5	2,462	3.2
2009	25,455	25,578	0.6	268	130.8	8.1	2,165	3.2

Table 2: Panel A provides yearly summary statistics for the S&P500 ETF (SPY). Panel B provides yearly summary statistics for the Dow Jones ETF (DIA). The variables are defined as follows: $\sigma(\Delta\text{time})$ is the standard deviation of the number of seconds between trades. σ(Size) is the standard deviation of the size of an individual trade. σ(Dursize) is the standard deviation of the size of a trade "string", where a "string" is defined as a group of consecutive buy transactions or sell transactions.

Panel A: SPY

Year	$\sigma(\Delta$ time)	σ(Size)	σ(Dursize)
1997	83.1	22959.3	39504.4
1998	31.1	21654.8	34738.1
1999	27.3	21456.5	32487.0
2000	25.9	21952.6	32737.4
2001	12.4	13433.3	19076.8
2002	4.2	11520.2	16777.2
2003	3.3	11724.2	19918.7
2004	2.5	7228.6	19152.8
2005	1.4	6918.0	21009.2
2006	1.4	8071.9	22622.1
2007	1.0	8370.3	27187.4
2008	0.4	6473.9	18897.4
2009	0.3	5103.2	16812.6

Panel B: DIA

Year	$\sigma(\Delta$ time)	σ(Size)	σ(Dursize)
1998	161.5	8150.1	13488.4
1999	117.1	10113.9	16118.8
2000	65.5	8794.0	13931.4
2001	35.4	8248.0	12147.4
2002	13.6	6706.9	9949.3
2003	9.0	4148.5	7166.9
2004	9.9	5521.1	11312.0
2005	9.0	4189.3	10762.5
2006	7.1	2683.5	8073.9
2007	4.9	3291.8	10967.3
2008	2.6	1968.6	5833.9
2009	2.5	1447.5	4974.4