

Exchange Volume Shares

Teddy Cho

August 10, 2015

1 Data

We use historical data from the NYSE Trade and Quote database¹, made available by Wharton Research Data Services. The dataset “contains intraday transactions data (trades and quotes) for all securities listed on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), as well as Nasdaq National Market System (NMS) and SmallCap issues”³.

Our data cover all trades for our sample of symbols⁴ over the period of Jan 1, 2014 - Jan 31, 2014. No time filtering is used, and we are left with 20 trading days total.

Future notes: Also explore how market shares look just before close auction, at different times of day, etc. Why do spikes happen in certain exchange. Look at times where we cannot explain, like institutional things. Consider throwing out expiration days, holidays, memorial day weekend.

2 Partitioning Data

2.1 Goals

From trades data, we wish to get a sense of exchange share over time. In order to transform trades data into exchange share data, we explore two methods of partitioning trades. The exchange shares are computed for each partition of trades.

¹We note that this data is known to have inaccuracies regarding quotes’ timestamps, and such issues may carry over to trade timestamps. These issues stem from timestamps originating from a secondary source (i.e., not the exchanges’ order matching engines) and being derived after an exchange aggregation method which introduces further latency².

³See <https://wrds-web.wharton.upenn.edu/wrds/ds/taq/index.cfm>.

⁴Our sample of symbols consists of AMD, BAC, BRKA, BRKB, C, GOOG, GRPN, JBLU, MSFT, RAD, SPY.

2.2 Time-Based

Time is partitioned into windows of equal length, and each trade is assigned to the partition which contains its timestamp. This method allowed for partitions to contain no trades, and we explored three ways to deal with these empty partitions.

Within a given symbol, the exchange i 's share for the time interval beginning at T is given by

$$p_{iT} = \frac{\sum_{T \leq t < (T+\Delta T)} q_{it}}{\sum_{T \leq t < (T+\Delta T)} q_t}$$

where q refers to a trade's volume and ΔT refers to the time interval.

2.2.1 Passing Over

Empty partitions are kept in determining lags, but excluded from contributing to correlation calculations. This is an established R option used via the `na.pass` argument⁵.

2.2.2 Discarding

Empty partitions are discarded before computing autocorrelations. For example, if partition t and window $t + n$ are separated by $(n - 1)$ empty partitions, they will be considered t and $t + 1$ after discarding the empty partitions.

2.2.3 Persisting Last State Through

To simulate that a regime simply continued if no trading activity occurred, we explored persisting a partition's exchange share through its trailing empty partitions (if they existed). This avenue has been closed, due to its positive biasing of autocorrelation.

2.3 Trade-Based

Trades are partitioned into sequential groups of n trades. To get a sense of the duration of these partitions, we compute a moving average of these partitions for our dataset.

⁵<https://stat.ethz.ch/R-manual/R-devel/library/stats/html/acf.html>

If we index volumes q of trades by k , the exchange i 's share at interval j is now given by

$$p_{ij} = \frac{\sum_{n \cdot j \leq k < n \cdot (j+1)} q_{ik}}{\sum_{n \cdot j \leq k < n \cdot (j+1)} q_k}.$$

3 Data Transformation

The main concern about using compositional data is the constraint that all the values sum to 1 in all periods.

Further, we wish to normalize an exchange's share based on the exchange's "typical" exchange share. To do this, we took the exchange x 's share p of stock s in the interval tp_{sxt} and used its deviation from the exchange x 's average share \bar{p} of stock s over the available timerange as a proportion of the average share. We call this z and use

$$z_{sxt} = \frac{p_{sxt} - \bar{p}_{sx}}{\bar{p}_{sx}}.$$

4 Autocorrelation

4.1 Goals

If there exist regimes during which certain exchanges are preferred, then we would expect positive autocorrelations within exchange.

4.2 Comparing Different Partitioning Methods

Consistently across symbols and intervals, we observe that trade-partitioning yields a clearer decay (and steeper decline) than the time-partitioning methods. Further, the figures reveal that when there is a low rate of empty intervals, the autocorrelation plots are very similar between discarding and passing over empty intervals.

As an example, we look to the dataset consisting of AMD trades on BATS EDGX.

One interesting thing is the persistence of positive autocorrelations far into the higher-order lags. I suspect this is due to the high number of (0,0) points which bunch up at the lower left corner of Quadrant I. Thus it is very hard for a negative correlation to overcome this "anchor". Maybe we should weight points on volume or trade count.

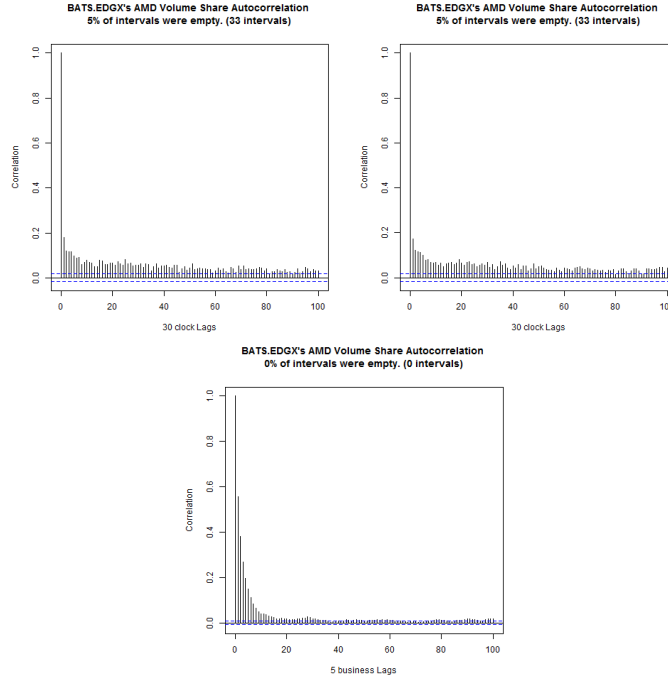


Figure 1: Autocorrelation plots for BATS EDGX’s exchange share: the top two plots used a time-partitioning scheme whereby trades were assigned to windows resulting from partitioning the timerange into 30 second intervals. The left plot passed over empty intervals as described in section 2.2.1, and the right plot discarded empty intervals as described in section 2.2.2. The third plot used a trade-partitioning scheme whereby trades were partitioned sequentially into groups of 5; note that this yields a more geometric decay and a steeper decline than the previous autocorrelation plots.

5 Inter-Exchange Correlations

We explore correlations among exchange shares within the same period, and correlations among exchange share across periods separated by a lag.

5.1 Within-Period Correlation

The following are correlation matrices for various symbols which embody trends representative of the sample.

The exchanges are ordered such that the minor exchanges⁶ are grouped at

⁶NSX, CBSX, NASDAQ.PSX, and CHX.

the top-left. They are followed by the Taker/Maker exchanges⁷. The final five exchanges are the Maker/Taker exchanges⁸.

The value in cell (i, j) is the correlation of exchange i 's share score with exchange j 's share score.

$$r_{z_i z_j} = \frac{\sum_{t=1}^n z_{it} z_{jt} - n \bar{z}_i \bar{z}_j}{(n-1) s_{z_i} s_{z_j}}$$

where z_{it} and z_{jt} are exchange volume share scores for exchange a and b at time t , and s_{z_i} and s_{z_j} are the sample standard deviations for z_{it} and z_{jt} , respectively.

⁷BATS.BYX, BATS.EDGA, and NASDAQ.BX

⁸NYSE, NYSE.Arca, BATS.BZX, and NASDAQ.

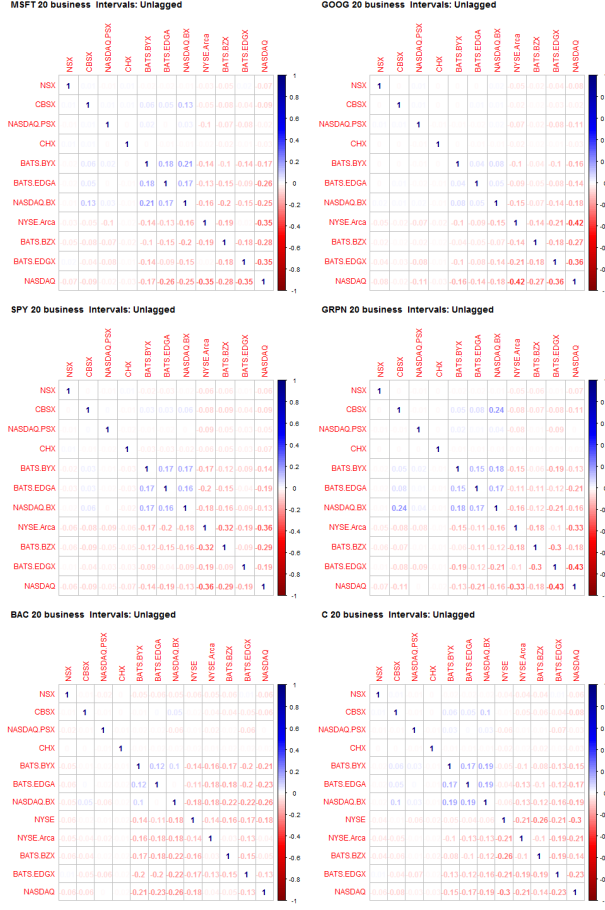


Figure 2: Correlation matrices for exchange shares in various stocks, using the trade-partitioning method for 20 trade intervals. Note correlations among the Taker/Maker exchange shares are positive. However, the Maker/Taker exchange shares show negative correlations between each other and against the Taker/Maker exchanges. There is an interesting pattern in BAC where the strongest correlations are those between exchanges of different fee structure; within the two groups of exchanges, correlations are less negative.

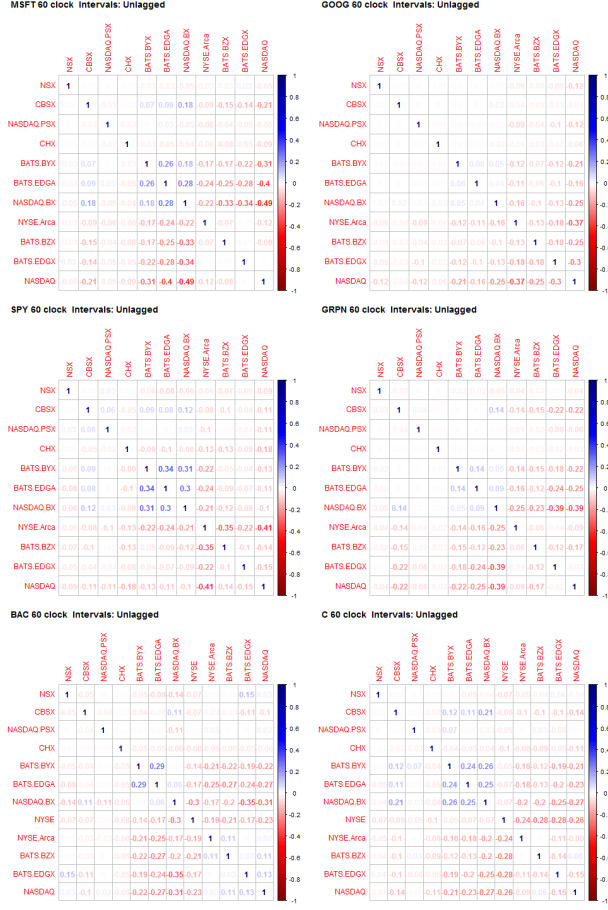


Figure 3: Correlation matrices for exchange shares in various stocks, using the time-partitioning method for 60 second intervals (method for dealing with empty intervals does not affect within-period correlations). The same patterns as in the previous set of graphs are present, with the particular correlations among Maker/Taker exchanges particularly dampened.

5.2 Single Lagged Correlation

The following are correlation matrices for various symbols which embody trends representative of the sample. The exchanges are ordered just as in section 5.1.

The value in cell (i, j) is the correlation of exchange i 's share with exchange j 's share in the next (lag 1) period.

$$r_{z_i z_j} = \frac{\sum_{t=1}^{n-1} z_{it} z_{j(t+1)} - n \bar{z}_i \bar{z}_j}{(n-2) s_{z_i} s_{z_j}}$$

where $z_{i,t}$ and $z_{j,t}$ are exchange volume share scores for exchange a and b at time t , and s_{z_i} and s_{z_j} are the sample standard deviations for z_{it} and z_{jt} , respectively.

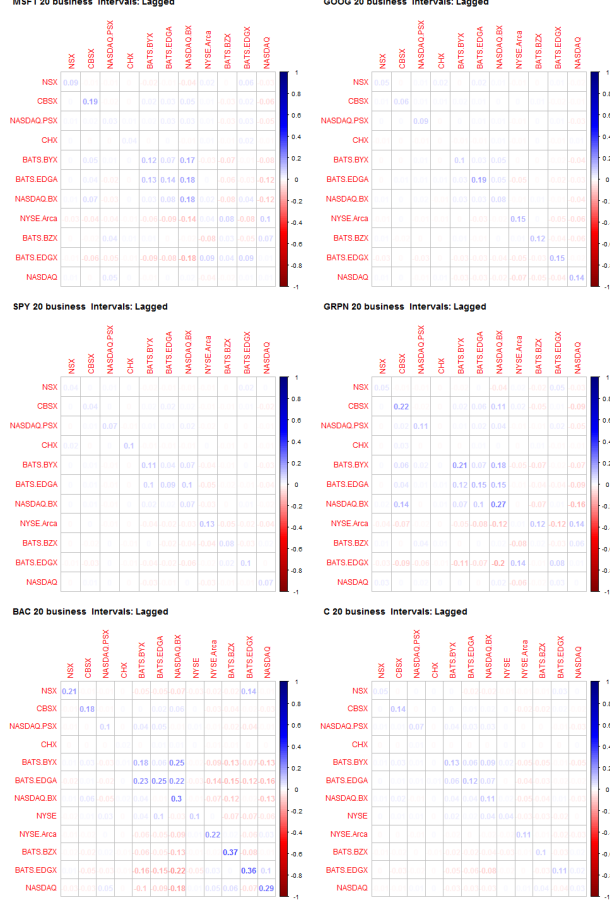


Figure 4: Correlation matrices for exchange shares in various stocks, using the trade-partitioning method for 20 trade intervals. The diagonal is simply the lag 1 autocorrelation for the respective exchange. Though weak, we can see a fainter inter-exchange effect analogous to within-period inter-exchange correlations.

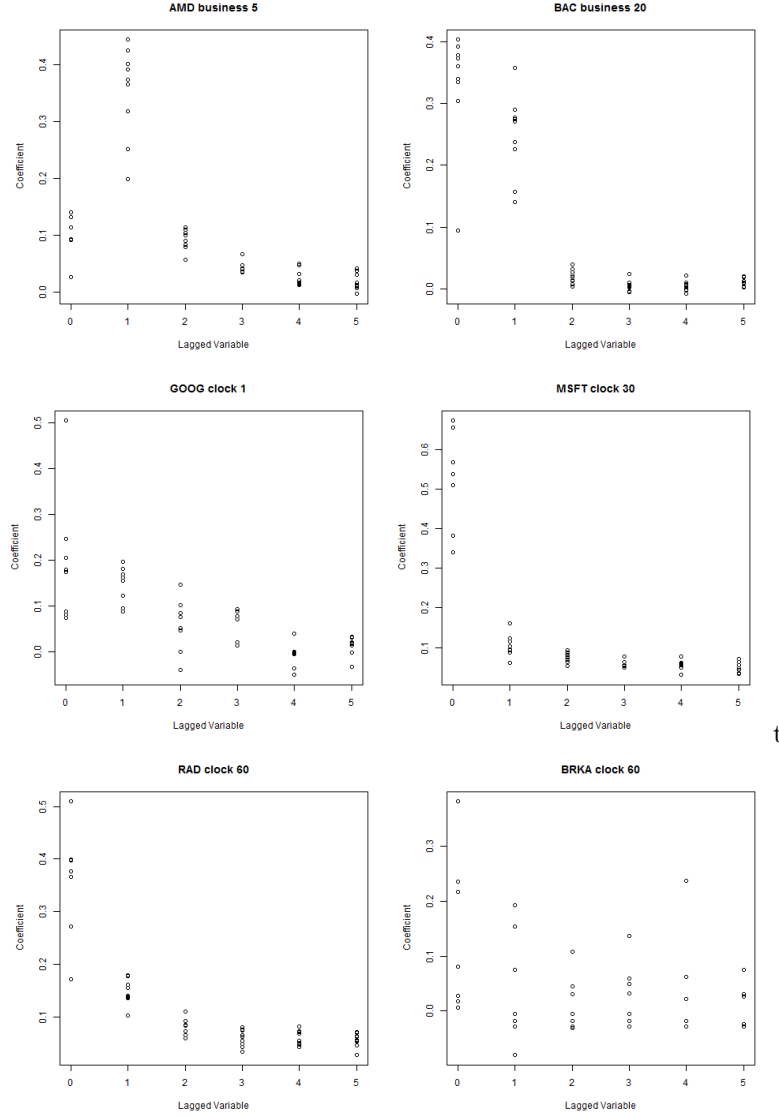


Figure 6: For a given symbol and partitioning method, each graph plots the regression coefficients for the intercept (on $x=0$) and for the lagged variables (on $x \in \{1, 2, 3, 4, 5\}$ for each respective lag). As each graph plots for all major exchanges, there is one coefficient for each exchange for each variable.

7 Variance

Variance of hourly market share, low intraday variance, high interday variance.

8 Compositional Data

Statistician Aitchison recommends logratios⁹.

⁹A Concise Guide to Compositional Data Analysis, 2003